INTERIM REPORT

2010 ESTCP UXO Classification Study, Camp Butner, NC

ESTCP Project MR-200910

April 2012

Dean Keiswetter
SAIC

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **APR 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**
4. TITLE AND SUBTITLE **2010 ESTCP UXO Classification Study, Camp Butner, NC**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SAIC,1710 SAIC Drive,McLean,VA,22102**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**151**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Contents

Figures

Tables

Acronyms

COV	Coefficient of Variation
DGM	Digital Geophysical Mapping
EE/CA	Engineering Evaluation/Cost Analysis
EMI	Electromagnetic Induction
ESTCP	Environmental Security Technology Certification Program
FAR	False Alarm Rate
FUDS	Formerly Used Defense Site
GLRT	Generalized Likelihood Ratio Test
GPS	Global Positioning System
IDA	Institute for Defense Analyses
IDL	Interactive Data Language
MM	Metal Mapper
MRS	Munitions Response Site
NOSLN	No On-Site Learning Necessary
NRL	Naval Research Laboratory
P_{class}	Probability of Correct Classification
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
QA	Quality Assurance
QC	Quality Control
ROC	Receiver Operating Characteristic
Rx	Receiver
SERDP	Strategic Environmental Research and Development Program
SI	Site Investigation
TEMTADS	Time Domain EM Discrimination Array
TOI	Targets of Interest
Tx	Transmitter
UXA	UX-Analyze
UXO	Unexploded Ordnance

Acknowledgements

This work was funded by the Environmental Security Technology Certification Program Office, under projects MM-0210 and MM-0901. The work was performed by a team of personnel from Science Applications International Corporation (SAIC), NAEVA and Parsons.

We gratefully acknowledge Dr. Herb Nelson, Program Manager of ESTCP's Munitions Management Thrust Area, Dr. Anne Andrews, SERDP and ESTCP Deputy Director, and Dr. Jeff Marqusee, SERDP and ESTCP Executive Director, for their vision, support, and guidance throughout this demonstration.

The Principal Investigator for this project was Dr. Dean Keiswetter, SAIC. Key technical contributors from SAIC include Mr. Tom Furuya, Dr. Jim Kingdon, Mr. Jonathan Miller, and Dr. Tom Bell. Key technical contributors from NAEVA include Alison Paski and Cora Blits. Key technical contributors from Parsons include Craig Murray, John Baptiste, and Steve Saville.

Executive Summary

The objective of this ESTCP-led UXO classification study was to demonstrate a methodology for the use of classification in the munitions response process. The three key components of this methodology were collection of high-quality geophysical data and principled selection of anomalous regions in those data, analysis of the selected anomalies using physics-based models to extract target parameters such as size, shape, and materials properties, and the use of those parameters to construct a prioritized dig list. Each of these components was handled separately in this program with SAIC focusing on the last two objectives. Under subcontract to SAIC, Parsons and NAEVA also analyzed and classified the selected anomalies.

SAIC's specific objective was to discriminate targets of interest from native clutter over an 4.5 acre area at former Camp Butner, North Carolina, by characterizing and classifying anomalies identified in electromagnetic induction (EMI) survey data. Both dynamic and cued (static) data were collected over the area. Misclassifying a target of interest as an item that can be left in the ground (viz., a false negative) was defined to be the primary failure. At this site, there were three main targets-of-interest (TOI): 37-mm, M48 Fuse and 105mm projectiles. No other types of UXO were found during the excavation phase of the demonstration.

The ESTCP Program Office coordinated data collection activities. This included all preparatory activities, arranging for a data collection by well-validated systems, selection of anomalies for analysis from each geophysical data set, and compilation of the individual sensor anomaly lists into a master list. Anomalies were selected from each geophysical data set using a target response-based threshold. Specifically, all anomalies corresponding to a 37mm projectile at 30cm depth (5.2mV in gate 2 for the EM61-MK2 cart were flagged for classification analysis. Validation digging was also coordinated by the Program Office.

The data analysis was done using UX-Analyze (an analysis framework within Oasis montaj™ that integrates quantitative analysis algorithms and custom-designed visualization schemes) and custom Interactive Data Language (IDL) inversion and processing routines. Specifically, SAIC, NAEVA and Parsons used UX-Analyze to process all anomalies identified by the Program Office from survey data collected using the EM61-MK2 cart, TEMTADS and Metal Mapper data. The cued TEMTADS and Metal Mapper data were also processed by SAIC using in-house custom IDL inversion and processing routines.

The data analysts were given the option to request a standard training set, a custom training set or none. These data were used as inputs to finalize classification algorithms and adjust thresholds. In the case of the no training data option, the thresholds were based on data from previous testing and training pit measurements. We tested all three types of training data in this demonstration. All analysis methods used a rules based classification approach with no "human veto" or interactive classification. At the conclusion of training, SAIC, NAEVA and Parsons submitted 10, 3, and 14 prioritized dig lists, respectively. The dig lists comprised of analyzing the following data: 1) EM61 only, 2) TEMTADS only, 3) MM only, 4) EM61 with TEMTADS, and 5) EM61 with MM. These lists were ordered from the item that is most confident not hazardous

(Category 1) through the item that is most confident munitions (Category 3). The anomalies in the transition zone between Category 1 and 3 for which we were not able to make a decision were assigned Category 2. The anomalies for which we were not able to extract meaningful parameters (Category 4) were placed at the bottom of the list. Only Category 1 anomalies were recommended to be left in the ground. These inputs were scored by the Institute for Defense Analyses (IDA) with emphasis on the number of items that were correctly labeled non-hazardous while correctly labeling all munitions items.

In general, the analysis methods that used the dynamically collected EM61-MK2 data either alone or as a pre-screener did not perform as well at this site. The EM61 data were only able to remove about 15% of the clutter items and still produced a false negative. Using the EM61 as a pre-screener reduced the number of cued targets by about 15-20% but also resulted in one or two false negatives. The main reason for the poor performance was that the target features (decay ratios and inverted target size) for the clutter and the TOI had significant overlaps at this site.

In contrast, the cued fixed array systems consisting of the TEMTADS and Metal Mapper produced polarizations that were accurate enough to discriminate between TOI and non-TOI on the basis of shape. Classification for the TEMTADS and Metal Mapper datasets were primarily based on an algorithm which compared our derived polarizabilities with a library of known target signatures. The ROC curves are similar for the TEMTADS and Metal Mapper although the TEMTADS' curve tends to start more vertical than the Metal Mapper's. Both of the cued sensors had ROC curves that were much better than those for the EM61 sensor. For most of the analysis methods using only the cued data, there were few false positives from the beginning of the curve until you reach about 95% of TOI recovered. Recovering the remaining 5% of TOI required a large number of unnecessary digs for most of the dig lists and at the user defined thresholds they typically resulted in single digit false negatives.

The false negatives from the cued systems were caused, in general, by unrecognized data problems or by inaccurate assumptions regarding the number of source objects. Data issues occurred for both the TEMTADS and the Metal Mapper sensors. For the TEMTADS sensor it was discovered during failure analysis that the polarity of coil 5 was reversed. We also discovered during failure analysis that the Geometrics MM system had intermittent problems with the Y axis coil for receivers 2 and 3. These data problems caused the false negatives. After fixing the data problems and using modestly different decision rules the TEMTADS data were able declare approximately 90% of the clutter as "Do not Dig" with no false negatives.

1.0 INTRODUCTION

This was the third in the series of ESTCP demonstrations of classification technologies for Munitions Response. This demonstration was designed to investigate the evolving classification methodology at a site contaminated with 37-mm projectiles in addition to larger munitions.

2.0 TECHNOLOGY

This demonstration used physics-based inversion algorithms and rule-based classification schemes to classify buried sources as targets of interest (TOI) or not. The analyses utilized data collected by a number of different Electromagnetic Induction (EMI) sensors systems as detailed below.

2.1 GEOPHYSICAL DATA COLLECTION

SAIC analyzed data acquired using an EM61-MK2 sensor, the Naval Research laboratory (NRL) Time Domain EM Discrimination Array (TEMTADS), and the MetalMapper (MM) system.

2.2 DATA ANALYSIS

Our IDL and UX-Analyze inversion algorithms assume a dipolar source and derive the best set of induced dipole model parameters that account for the spatial variation of the signal as the sensor is moved over the object. The model parameters are target X,Y,Z location, three dipole response coefficients corresponding to the principle axes of the target, and the three angles that describe the orientation of the dipole. We utilized single-source and multi-source solvers.

2.3 CLASSIFICATION

For the TEMTADS and MM sensor data, our library match algorithm compares the polarizabilities of an unknown target with each library entry based on 3 criteria: the amplitude of the principal polarization ($\beta 1$), and the two shape parameters, which are ratios of the second- or third-principal polarization to the first ($\beta 2/\beta 1$ or $\beta 3/\beta 1$; respectively). The difference in the values is computed at all time gates, excluding those where the values are negative. The results from the 3 different criteria were averaged, producing a metric which ranges from 0 (worst possible fit) to 1 (perfect fit).

The classification method for EM61 utilized peak amplitude, spatial footprint, fitted size, and decay ratios.

3.0 PERFORMANCE OBJECTIVES

The performance objectives for this demonstration are summarized in Table 3-1 to Table 3-15. We have included performance objectives for SAIC's analysis using UX-Analyze and IDL based analysis routines as well as analyses performed by NAEVA and Parsons using UX-Analyze. Since this is a classification demonstration, the performance objectives focus on target analysis and classification; we assume that the anomalies from all targets of interest have been detected and included on the target list the analysis demonstrators worked from.

The first three objectives refer to the classification part of the demonstration with the first two referring to the best results from each approach in a retrospective analysis and the third addressing how well each demonstrator is able to specify the correct threshold in advance. The final two objectives refer to the feature extraction part of the demonstration. Many of the different classification methods use the same target parameters therefore we have only included these parameters in one of the pertinent tables.

Table 3-1 Performance Objectives for EM61 MK2 Cart - UXA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	170 of 171 (99.4%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	119 of 2119 (5.6%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = .994 N_{fa} = 1801
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	9 of 2290 (0.4%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=1.0,0.74,0.87 XY mean = 17cm (11cm σ) Z mean= -15cm (14cm σ) Size COV=0.237

Table 3-2 Performance Objectives for EM61 MK2 Cart with TEMTADS - UXA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	162 of 171 (94.7%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	0 of 2119 (0%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 0.947$ $N_{fa} = 977$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as “Unable to Analyze.”	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor’s detection list.	0 of 2290 (0.0%) classified as “Unable to analyze”

Table 3-3 Performance Objectives for EM61 MK2 Cart with Metal Mapper – UXA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	163 of 171 (95.3%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	0 of 2119 (0%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 0.953$ $N_{fa} = 1337$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as “Unable to Analyze.”	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor’s detection list.	16 of 2290 (0.70%) classified as “Unable to analyze”

Table 3-4 Performance Objectives for TEMTADS - UXA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	168 of 171 (98.2%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	432 of 2119 (20.4%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.982 N_{fa} = 1242
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	0 of 2290 (0.0%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.21,0.17,0.18 XY mean = 10cm (5cm σ) Z mean= 3cm (5cm σ) Size COV=0.06

Table 3-5 Performance Objectives for TEMTADS – 2 criteria - IDL

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	164 of 171 (95.9%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	407 of 2119 (19.2%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.959 N_{fa} = 533
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	2of 2290 (0.1%) classified as "Unable to analyze"

Table 3-6 Performance Objectives for TEMTADS – 3 criteria - IDL

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	164 of 171 (95.9%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	285 of 2119 (13.4%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.959 N_{fa} = 526
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	2of 2290 (0.1%) classified as "Unable to analyze"

Table 3-7 Performance Objectives for TEMTADS – NOSLN - IDL

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	164 of 171 (95.9%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	263 of 2119 (12.4%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.959 N_{fa} = 543
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	2of 2290 (0.1%) classified as "Unable to analyze"
Correct estimation of TOI target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs@0.3ms :COV<0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size@0.3ms:COV<0.20	β1,2,3 COV=0.22,0.19,0.22 XY mean = 12cm (7cm σ) Z mean= 3cm (10cm σ) Size COV=0.07

Table 3-8 Performance Objectives for Metal Mapper – 2 criteria - IDL

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	171 of 171 (100%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	301 of 2119 (14.2%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 1.0 N_{fa} = 1882
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	44of 2290 (1.9%) classified as "Unable to analyze"

Table 3-9 Performance Objectives for Metal Mapper – 3 criteria - IDL

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	171 of 171 (100%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	332 of 2119 (15.6%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 1.0 N_{fa} = 1824
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	44of 2290 (1.9%) classified as "Unable to analyze"

Table 3-10 Performance Objectives for MetalMapper – NOSLN

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	164 of 171 (95.9%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	785 of 2119 (37%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator -specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 0.959$ $N_{fa} = 1009$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	44 of 2290 (1.9%) classified as "Unable to analyze"
Correct estimation of TOI target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs@0.3ms :COV<0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size@0.3ms:COV<0.20	β1,2,3 COV=0.38,0.46,0.63 XY mean = 12cm (7cm σ) Z mean= 3cm (10cm σ) Size COV=0.10

Table 3-11 Performance Objectives for MetalMapper – 3criteria -NAEVA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	170 of 171 (99.4%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	1092 of 2119 (52%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 0.994$ $N_{fa} = 777$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	363 of 2290 (15.8%) classified as "Unable to analyze"

Table 3-12 Performance Objectives for MetalMapper – 3criteria/1 criteria -NAEVA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	171of 171 (100%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	1313 of 2119 (62%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 1$ $N_{fa} = 1016$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	343 of 2290 (15.0%) classified as "Unable to analyze"

Table 3-13 Performance Objectives for MetalMapper – 3criteria/2 criteria -NAEVA

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	171of 171 (100%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	1252 of 2119 (64%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 1$ $N_{fa} = 1005$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	346 of 2290 (15.1%) classified as "Unable to analyze"

Table 3-14 Performance Objectives for EM61 MK2 Cart - Parsons

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	127 of 171 (74%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	247 of 2119 (11.6%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	$P_{class} = 0.74$ $N_{fa} = 871$
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	272 of 2290 (11.9%) classified as "Unable to analyze"

Table 3-15 Performance Objectives for EM61 MK2 Cart and MetalMapper - Parsons

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	166 of 171 (97.1%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	255 of 2119 (12%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.971 N_{fa} = 693
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	263 of 2290 (11.5%) classified as "Unable to analyze"

3.1 MAXIMIZE CORRECT CLASSIFICATION OF MUNITIONS

This is one of the two primary measures of the effectiveness of this approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to be able to classify the targets with high efficiency. This objective concerns with the component of the classification problem that involves correct classification of items-of-interest.

3.1.1 Metric

The metric for this objective is the number of items on the master anomaly list that can be correctly classified as munitions by each classification approach.

3.1.2 Data Requirements

A prioritized dig list for the targets on the master anomaly list was prepared for each of the data sets analyzed as part of this demonstration. Using complete ground truth of all the anomalies on the master list IDA personnel used their scoring algorithms to assess the results.

3.1.3 Success Criteria

The objective will be considered to be met if all of the items-of-interest are correctly labeled as munitions on the prioritized anomaly list.

The probability of correct classificaiton ranged from 74% to 100% among the different classification schemes with the vast majority exceeding 95%. The classification scheme that did the poorest used only the EM61-MK2 data and the large number of failures was due to an aggressive choice of the "dig/no dig" threshold which placed several TOIs in the no dig category. The classification schemes using only MM data had the best results with many of the lists either classifying all 171 ordnance items as TOI or having only one failure. The TEMTADS only lists also did well but had a few more false negatives. The classification schemes that used the EM61-Mk2 data as a pre-screener to limit the number of targets requiring MM or TEMTADS cued analysis did slightly worse than just using one of the advanced sensors but performed better than using only the EM61-MK2 data.

3.2 MAXIMIZE CORRECT CLASSIFICATION OF NON-MUNITIONS

This is the second of the two primary measures of the effectiveness of this approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to be able to classify the targets with high efficiency. This objective concerns false alarm reduction.

3.2.1 Metric

The metric for this objective is the number of items-of-interest on the master dig list that can be correctly classified as non-munitions by each classification approach.

3.2.2 Data Requirements

See section 3.1.2.

3.2.3 Success Criteria

The objective will be considered to be met if more than 30% of the non-munitions items can be correctly labeled as non-munitions while retaining all of the targets-of-interest on the dig list.

The lists generated using the EM61-MK2 data either alone or in conjunction with the MM or TEMTADS data did not meet the 30% success criteria. They ranged from correctly classifying none to 12% of the non-munitions items before reaching a TOI. The main issue was that the size and measured time decay of the clutter items calculated using the EM61-MK2 data overlapped significantly with the TOI (especially the 37mm) at this site. The lists generated using only the TEMTADS data also did not meet the success criteria. Overall the TEMTADS data analysis performed fairly well and all lists correctly classified 95% of the TOI with only about 50 false positives. Unfortunately, the remaining 5% were scattered throughout the dig lists so only 7%-20% of the non-munitions items were correctly classified before reaching a TOI. The MM data

analysis using SAIC's IDL based analysis software performed similar to the TEMTADS analysis with only 14%-21% of the non-munitions items correctly classified before reaching a TOI. On the other hand the MM analysis performed by NAEVA using UX-Analyze reached the objective by correctly classifying 52% to 64% of the non-munitions items before reaching a TOI.

3.3 SPECIFICATION OF NO-DIG THRESHOLD

In a retrospective analysis as was performed in this demonstration, it is possible to tell the true classification capabilities of a classification procedure based solely on the prioritized dig list submitted by each demonstrator. In a real-world scenario, all targets may not be dug so the success of the approach will depend on the ability of an analyst to accurately specify their dig/no-dig threshold.

3.3.1 Metric

P_{class} and number of false alarms, N_{fa}, at the demonstrator-specified threshold are the metrics for this objective.

3.3.2 Data Requirements

See section 3.1.2.

3.3.3 Success Criteria

The objective will be considered to be met if more than 30% of the non-munitions items can be correctly labeled as non-munitions while retaining all of the targets-of-interest at the demonstrator-specified threshold.

The only lists that met this objective were the MM 3criteria/1criteria and 3criteria/2criteria methods using UX-Analyze. These lists met the objective by classifying all 171 TOIs as anomalies to dig while correctly labeling 52% and 52.6% of the non-munitions at the chosen threshold, respectively. The MM 3criteria method using UX-Analyze came close to meeting the objective by classifying 170 of 171 TOIs as anomalies to dig while correctly labeling 63.3% as non-munitions. A couple of the lists generated using SAIC's IDL based analysis software on MM data also classified all 171 TOIs as anomalies to dig but only was able to correctly label 11.2% and 13.9% of the non-munitions at the chosen threshold. The EM61-MK2 cart list using UX-Analyze classified 170 of 171 TOI as anomalies to dig but only was able to correctly label 15% of the non-munitions at the chosen threshold. The remainder of the lists met the 30% criteria but also had several false negatives at the chosen threshold.

3.4 MINIMIZE NUMBER OF ANOMALIES THAT CANNOT BE ANALYZED

Anomalies for which reliable parameters cannot be estimated cannot be classified by the classifier. These anomalies must be placed in the dig category and reduce the effectiveness of the classification process.

3.4.1 Metric

The number of anomalies for which reliable parameters cannot be estimated is the metric for this objective.

3.4.2 Data Requirements

A list of all target parameters along with a list of those anomalies for which parameters could not be reliable estimated was submitted for each of the data sets analyzed as part of this demonstration.

3.4.3 Success Criteria

The objective will be considered to be met if reliable parameters can be estimated for > 90% of the anomalies on each sensor anomaly list.

The majority of the dig lists easily met this objective with target parameters calculated for 98-100% of the anomalies on each sensor anomaly list. The exceptions were the 3 versions based on MM only data using UX-Analyze which had approximately 15% targets classified as "can't analyze". The two lists generated by Parsons also failed to meet the criteria with about 12% of the targets classified as "can't analyze".

3.5 CORRECT ESTIMATION OF TARGET PARAMETERS

This objective involves the accuracy of the target parameters that are estimated in the first phase of the analysis. Successful classification is only possible if the input features are internally consistent. The obvious way to satisfy this condition is to estimate the various target parameters accurately.

3.5.1 Metric

Accuracy of estimation of target parameters is the metric for this objective. The metric used for the XY location and depth accuracy is the mean and standard deviation of difference between the estimated and true values. The metric used for the estimation of size and betas is the coefficient of variation (COV), which is simply the standard deviation divided by the mean value of the parameter. It essentially reports variability in the estimated parameter. We used this normalized metric to compare performance because different EMI sensors return polarization estimates that vary in absolute magnitude.

3.5.2 Data Requirements

A list of all target parameters was submitted for each of the data sets analyzed as part of this demonstration. SAIC analysts compared these estimated parameters for all TOI with good isolated or marginally overlapping signals.

3.5.3 Success Criteria

The objective will be considered to be met if the estimated βs are within ± 20%, the estimated X, Y locations are within 15 cm (1σ), the estimated depths are within 10 cm (1σ), and the estimated size is within ± 20%.

The EM61-MK2 sensor came close to meeting the target parameter objectives for size and XY location but results for Z and individual betas showed much more variability. The XY location had a mean of 17cm with a standard deviation of 11cm. The Z had a mean of -15cm with a standard deviation of 14cm. The size had a COV of 0.237 which slightly exceeded the goal of 0.2 but was much better than the COV of the individual betas which ranged from 0.74 to 1.0.

The cued array systems produced more accurate target parameters than the dynamic system. Both the TEMTADS and the MM met the objectives for size, XY location and Z but did not quite meet the objectives for the individual betas. The individual betas for the TEMTADS were close to the 0.20 criteria with those calculated using UX-Analyze actually meeting the criteria for Beta2 and Beta3. The individual betas for the MM using SAIC's IDL analysis software showed much more variability with COV's ranging from 0.38 to 0.63 for the individual betas.

4.0 SITE DESCRIPTION

The site for this demonstration is former Camp Butner, which is located approximately 15 miles north of Durham, North Carolina.

4.1 MUNITIONS CONTAMINATION

A large variety of munitions have been reported at the former Camp Butner. These include:

- Rifle grenades
- 2.36-inch rockets
- 37mms
- 40mms
- 81mm mortars
- 105mms
- 155mms
- 240mms

At the particular site of this demonstration, 105-mm and 37-mm have been observed. The excavation of two grids as part of the preparatory activities provided evidence of 37-mm projectiles and heavy-walled fragments.

4.2 SITE CONFIGURATION

The demonstration site was configured as four areas that total approximately 9.8 acres. The four demonstration areas are shown in Figure 4-1. Due to the number and density of anomalies only a portion of the northeast area was used to select anomalies that were given to the analysis teams for classification.

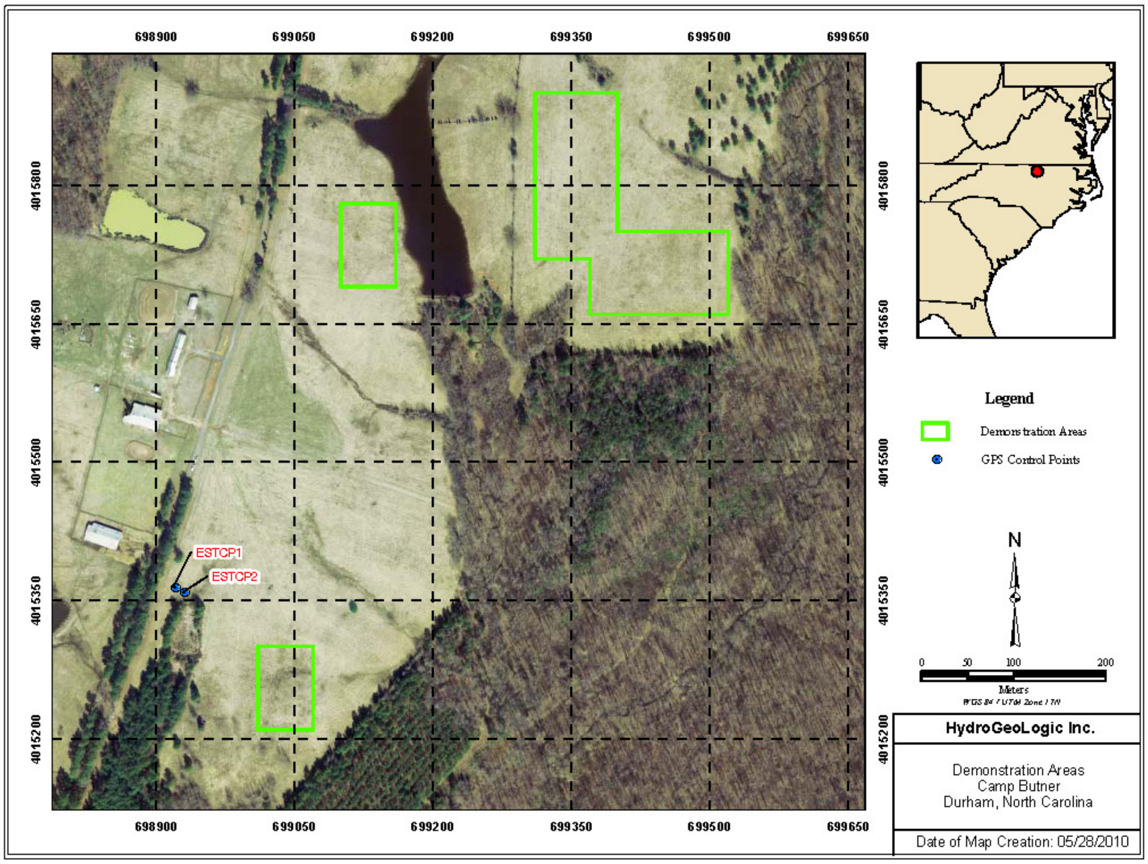


Figure 4-1. Final Camp Butner Demonstration Areas.

5.0 TEST DESIGN

5.1 CONCEPTUAL EXPERIMENTAL DESIGN

The objective of this program was to demonstrate a methodology for the use of classification in the munitions response process. The three key components of this methodology are collection of high-quality geophysical data and principled selection of anomalous regions in those data, analysis of the selected anomalies using physics-based models to extract target parameters such as size, shape, and materials properties, and the use of those parameters to construct a prioritized dig list. Each of these components was handled separately in this program.

The ESTCP Program Office coordinated data collection activities. This included all preparatory activities, arranging for a data collection by well-validated systems, selection of anomalies for analysis from each geophysical data set, and compilation of the individual sensor anomaly lists into a master list.

SAIC, and other data analysts, processed the individual data sets using existing routines to extract target parameters. These parameters were passed to the classification routines which, after training on a limited amount of site-specific ground truth, were used to produce prioritized dig lists.

Validation digging was coordinated by the Program Office. Since this was a demonstration, all anomalies on the master dig list were investigated. The underlying target was uncovered, photographed, located with a cm-level GPS system, and removed. The identities of a small number of the recovered items were provided to the demonstrators as training data if requested; and they used these inputs to finalize algorithms and adjust thresholds.

At the conclusion of training, SAIC and the other analysts submitted a prioritized dig list for each data set they have analyzed. These lists were ordered from the item the demonstrator is most confident is not hazardous through the item the demonstrator is most confident is a munitions. The anomalies for which the demonstrator was not able to extract meaningful parameters were placed at the bottom of the list. These inputs were scored by the Institute for Defense Analyses with emphasis on the number of items that were correctly labeled non-hazardous while correctly labeling all munitions items.

The primary objective of the demonstration was to assess how well each demonstrator was able to order their prioritized anomaly list and specify the threshold separating high confidence clutter from all other items. The secondary objective was to determine the classification performance that could be achieved by each approach through a retrospective analysis.

5.2 PRE-DEMONSTRATION ACTIVITIES

The ESTCP Program Office seeded the site with 37-mm and 105-mm projectiles and M48 fuze assemblies.

A quiet area near the entrance of the site was located to establish a sensor validation strip and a training pit. The training pit was used by the data collectors to measure sensor response for representative 37-mm and 105-mm projectiles and a M48 fuze assembly at a couple of different burial depths and orientations.

6.0 DATA ANALYSIS PLAN

SAIC submitted multiple dig lists. Each dig list possessed a unique combination of data type, training data, analysis firm, and processing environment (Table 6-1). All of our analyses utilized the standard intrinsic polarizabilities and extrinsic features of the dipole model. Additionally, all of the classification approaches utilized rule-based decision criteria.

Under subcontract to SAIC, Parsons and NAEVA used UX-Analyze to invert and classify EM61 and MM data. They consulted with SAIC regarding processing methodologies and decision thresholds, but coordinated directly with ESTCP for training label requests and dig list submittals.

Table 6-1. SAIC Data Analysis Plan

	Dig Lists	Sensor Data	Training Data	Features Used for Classification	Multiple Submittals
IDL	1	TEMTADS (all cued)	Existing library only (no onsite labels) [aka NOSLYN]	Polarizabilities	Yes
	2		Existing library + requested onsite labels	Two Polarizabilities (principal and secondary)	No
	3			Polarizabilities	
	4	MM (all cued)	Existing library only (no onsite labels) [aka NOSLYN]	Polarizabilities	Yes
	5		Existing library + requested onsite labels	Two Polarizabilities (principal and secondary)	No
	6			Polarizabilities	
UX-Analyze SAIC	7	EM61	Existing library + requested onsite labels	Estimated Size and measured decay	No
	8	EM61 + requested TEMTADS	Existing library only (no onsite labels) [aka NOSLYN]	Polarizabilities	
	9	TEMTADS (all cued)		Polarizabilities	
	10	EM61 + requested MM		Polarizabilities	
UX-Analyze Parsons	11	EM61	Existing library + requested onsite labels	Measured decay	No
	12	EM61 + requested MM	Existing library + requested onsite labels	Polarizabilities	
UX-Analyze NAEVA	13	MM (all cued)	Existing library + requested onsite labels	Polarizabilities	No
	14			Polarizabilities Two Polarizabilities (principal and secondary)	
	15			Polarizabilities Principal Polarizability	

6.1 PREPROCESSING

Survey data were preprocessed (located and simple filtering) by the data collection demonstrators in preparation for anomaly selection. Details of preprocessing steps can be found in the individual demonstrator demonstration plans [1-2]. Overall the data provided by the data collection demonstrator was deemed satisfactory and was used by SAIC for target parameter extraction.

6.2 PARAMETER ESTIMATION

Our discrimination approach uses a model-based estimation procedure to determine whether or not an unknown target is likely to be a UXO item. It entails estimating the size and shape of the target from the spatial pattern of the induced field above the target [3,4,5]. The EMI signal is a linear function of the flux through the receiving coil. In our model, the flux is assumed to originate from an induced dipole moment at the target location given by:

$$\mathbf{m} = \mathrm{UBU}^{\mathrm{T}}\mathbf{H}_0$$

where $\mathbf{H}_0$ is the peak primary field at the target, U is the transformation matrix between the coordinate directions and the principal axes of the target, and B is an empirically determined, effective magnetic polarizability matrix. For an arbitrary compact object, this matrix can be diagonalized about three primary body axes and written as:

$$\mathrm{B} = \begin{bmatrix} \beta_X & 0 & 0 \\ 0 & \beta_Y & 0 \\ 0 & 0 & \beta_Z \end{bmatrix}.$$

The relative magnitudes of the β's are determined by the size, shape and composition of the object as well as the transmit waveform and time gate or frequency. The transformation matrix contains the angular information about the orientation of these body axes.

For cylindrical objects like most UXO, B is a diagonal matrix with only two unique coefficients, corresponding to the longitudinal (β_T) and transverse (β_L) directions:

$$\mathrm{B} = \begin{bmatrix} \beta_L & 0 & 0 \\ 0 & \beta_T & 0 \\ 0 & 0 & \beta_T \end{bmatrix}.$$

Discrimination is based on target β's estimated from spatially mapped data. Specific ordnance items have specific β values, while clutter items generally have different β values.

6.3 TRAINING

All of the classification approaches required some level of training data. These data came from three sources:

- Sensor data for the targets-of-interest collected in previous testing,
- Data collected over the training pit, and
- Ground truth from excavations at the site

Each of the classification approaches used some combination of these data. The No On-Site Learning Necessary (NOSLN) approaches only used the first two sources of training data. The non-NOSLN IDL and Parson's approaches added a custom set of excavations whereas the approaches by NAEVA and UX-Analyze EM61-MK2 used the standard collection of digs.

6.4 CLASSIFICATION

SAIC used the above provided training data to train our algorithms and produced a ranked anomaly list for each of the sensor data sets that were processed. The list followed the format shown in Figure 6-1.

Rank	Anomaly ID	$P_{clutter}$	Comment	
1	247	.97		
2	1114	.96	High confidence NOT munition	
3	69	…		
…	…	…		Threshold
…	…	…		
…	…	…	Can't make a decision	
…	…	…		
…	…	…		
…	…	…		
…	…	…	High confidence munitions	
…	…	.03		
…	…	.02		
…	…			
N-2	…			
N-1	…		Can't extract reliable features	
N	…			

Figure 6-1. Format for the prioritized anomaly lists.

The first item on each anomaly list was that item which SAIC is the most confident is not a munition. The items were ranked according to decreasing confidence that the item is not hazardous. We provided two thresholds. The first threshold, or demonstrator operating point, is indicated in the figure and corresponds to the last item that can be classified as "high confidence not a munition." We will also indicate which item is the first that can be classified as "high confidence munition." All targets for which reliable parameters cannot be extracted must be dug and will be placed at the bottom of the list.

It is possible that groups of anomalies in the top three categories may have equal rank. Those in the "can't extract reliable features" group were listed in any order as there is no way to distinguish among them. If multiple items are associated with an entry on the list, the anomaly received one ranking appropriate to the most hazardous item in the collection.

The specific training and classification procedures for the data sets analyzed by SAIC are discussed individually in the sections that follow. The procedures used by NAEVA and Parsons are found in Appendices B and C.

6.4.1 EM61 Data

The EM61-MK2 cart data collected at Camp Butner were used in two different ways. First, we used the EM61 data as a pre-screener to determine which targets could be confidently classified as high confidence clutter or high confidence TOI thereby saving the costs associated with collecting cued data over the target. Second, we compiled a ranked dig list based solely on our analysis of the EM61 data. We used UX-Analyze to characterize and classify the EM61 data.

The EM61 pre-screener analysis only used the test pit measurements to develop rules to classify the targets. Figure 6-2 presents the inverted fit size versus the decay ratio as measured by a fourth to first time gate (1266 and 216 μsec, respectively) average for measurements near that targets' peak signal. The solid cyan, red and green squares show the test pit measurements of the 105mm, M48 fuze and 37mm, respectively. The solid black squares represent the unknown targets that have good data quality while the gray squares indicate unknown targets with data problems such as overlapping signatures that cause problems with the data inversions. The colored circles show the anomalies that possess features that match those for the different TOI and clutter items. As the figure shows, the size metric can be used to distinguish the 37mm and the M48 fuze from the 105mm rounds, but the 37mm and M48 fuze cannot be easily separated using fitted size metric. Also, the range of the size metric for all the TOI overlaps with virtually all the clutter items thus making the size metric ineffective for discriminating between TOI and clutter. On the other hand, the decay information does a better job of discriminating between the TOI and the non-TOI but it also has limited potential. Figure 6-3 presents the mean versus standard deviation of the fourth to first time gate data (channel 4 to channel 1). It uses the same symbols to represent the TOI but the black squares are anomalies that show good or decent data quality. We relaxed the data quality criteria when solely looking at the decay values because they are not as sensitive to data problems as the data inversions. We also analyzed the standard deviation of decay parameter in conjunction with the mean decay to ascertain whether multiple

objects of different compositions or wall thickness were present. If the standard deviation was greater than 0.03 we placed the anomaly in the "cued data needed" category even if the mean decay was below the 0.1 threshold because the higher standard deviation indicated the possibility of multiple objects.

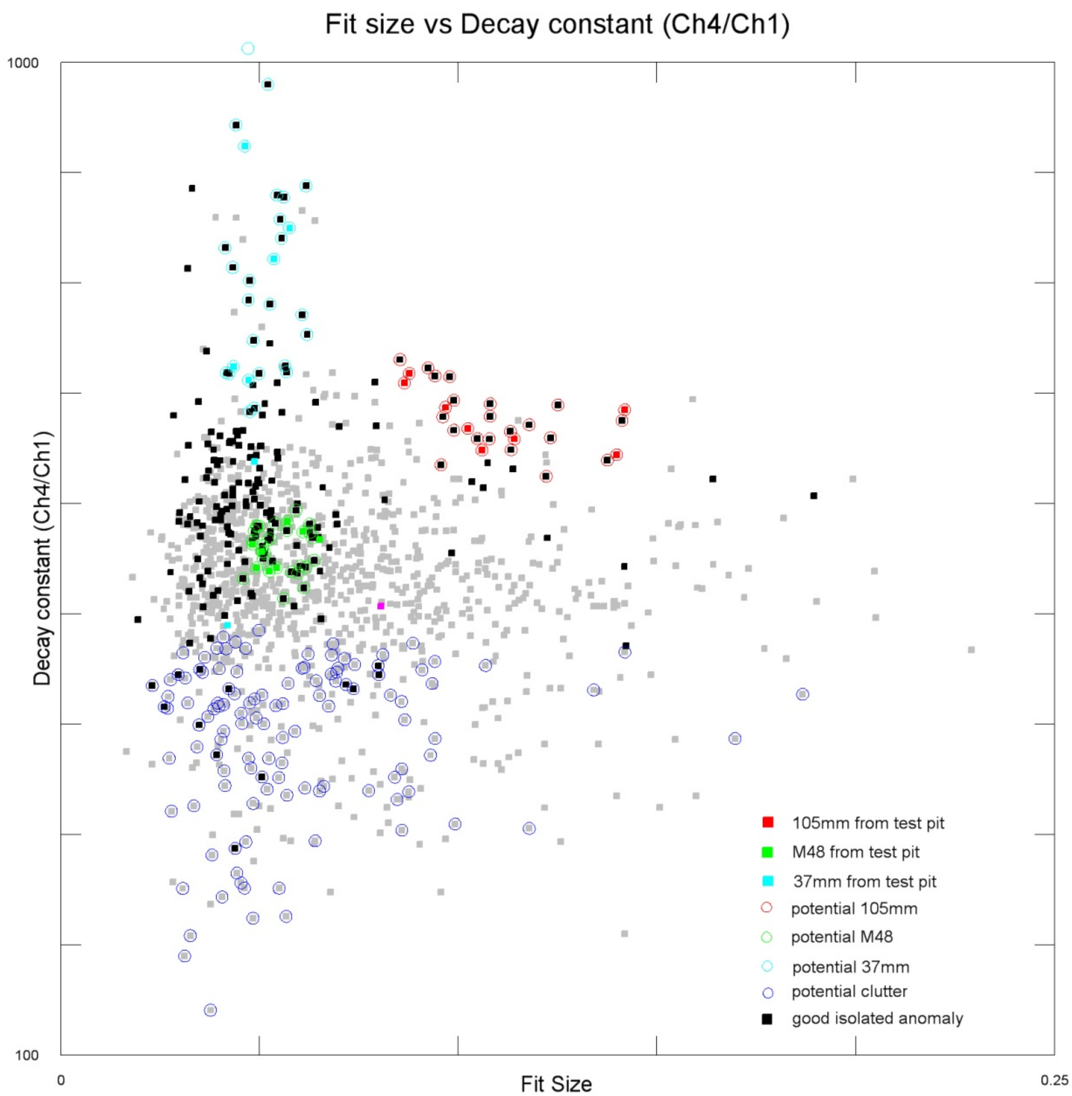


Figure 6-2. Inverted size parameter versus observed time decay for the test pit measurements and all the unknown anomalies, EM61 cart data.

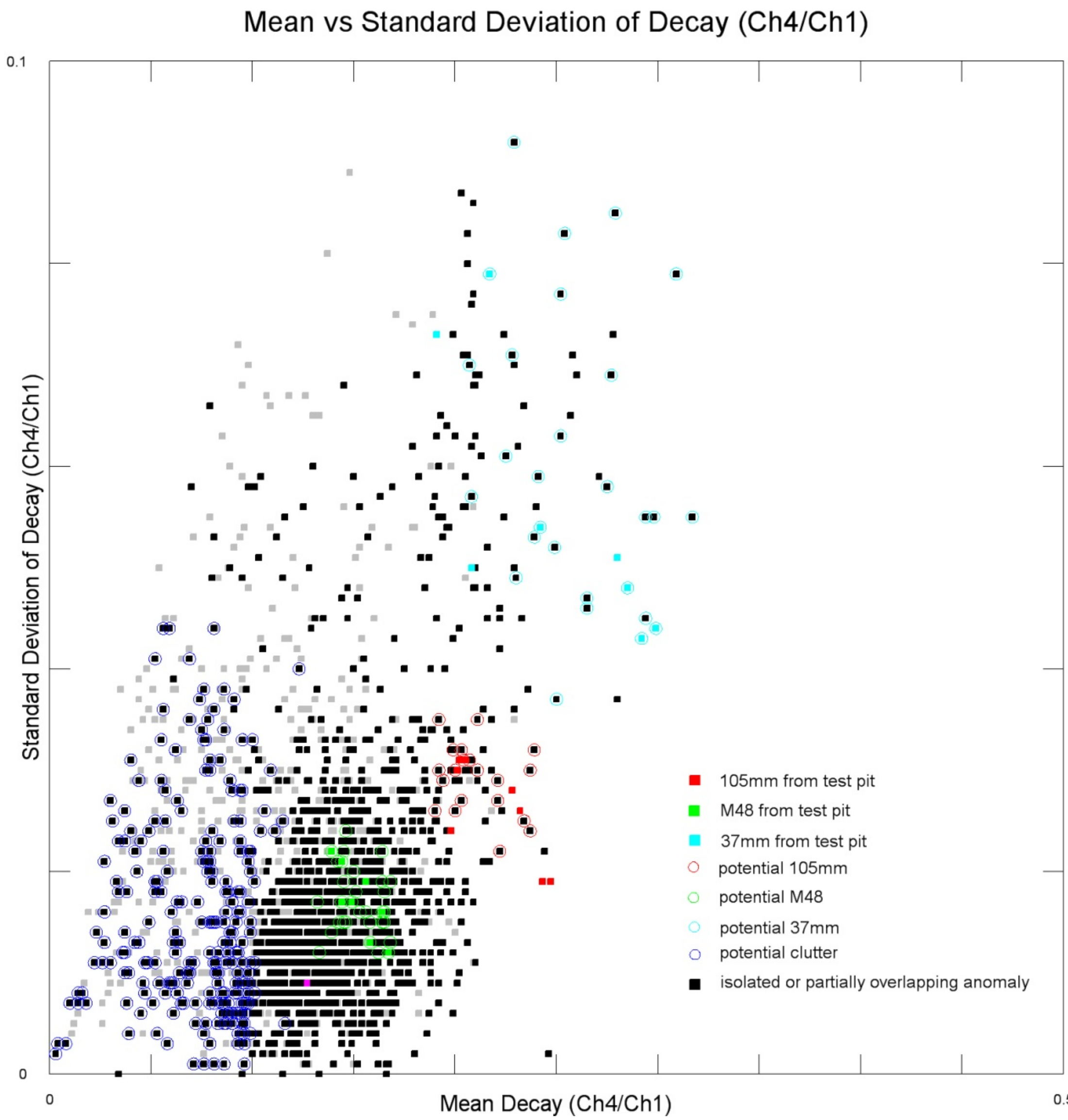


Figure 6-3. Standard deviation of Decay (Ch4/Ch1) versus Mean Decay (Ch4/Ch1) for the test pit measurements and all the unknown anomalies, EM61 cart data.

The EM61 prescreening decision was a straight-forward, multi-step process. To find the high confidence TOI anomalies, we used the mean and standard deviation of the extracted decay rates to calculate the Generalized Likelihood Ratio Test (GLRT) probability that the test item was each of the TOI types individually and kept the target type that produced the highest confidence value. If the data were of good quality, the fit error was low and the fitted size was within a reasonable range (see Table 6-2) determined by the test pit measurements, we flagged the anomaly as high confidence TOI. Because this was a pre-screener we selected ranges that were

tighter than normal to increase the likelihood of selecting a TOI. Next, we calculated the final decision statistic by combining the GLRT probability of the extracted decay rates and fitted size for the individual TOI type.

Table 6-2. Range of fitted size for TOI.

TOI	Fit Size range (m)
37mm	.039 to .060
M48 Fuze	.045 to .065
105mm	.085 to .125

To determine the high confidence clutter anomalies, we set thresholds for different decay based target parameters. The parameters consisted of the mean and standard deviation of the measured decay using Channels 1 and 4. We also used the TAU calculation found in UX-Analyze that is also a decay base parameter that compares Channels 1 to Channels 3 and 4 (TAU14 and TAU13). If the data were of good quality and the anomaly met the clutter criteria for two or more of the target parameters in Table 6-3 it was classified as a high confidence clutter.

Table 6-3. Thresholds for determination of high confidence clutter.

Target parameter	Range
TAU Ch4/Ch1	<450
TAU Ch3/Ch1	<325
Mean Decay Ch4/Ch1	<0.1
Standard deviation of Decay Ch4/Ch1	<0.03

The targets in Figure 6-2 and Figure 6-3 that are circled show the final results of the EM61 pre-screener. 319 out of 2290 anomalies were classified using the EM61 prescreening techniques with the remaining anomalies requiring cued data to make the final classification.

The EM61 only analysis used the standard collection of digs as well as test pit measurements to develop rules to classify the targets. Figure 6-4 presents the inverted fit size versus the decay ratio as measured by a fourth to first time gate (1266 and 216 µsec, respectively) average for measurements near that targets' peak signal. Similar to the previous figures, the solid cyan, red and green squares show the training data of the 105mm, M48 fuze and 37mm, respectively. The solid black squares represent the training targets from the standard training set that have good data quality while the gray squares indicate training targets with data problems such as overlapping signatures that cause problems with the data inversions. The results are similar to the EM61 pre-screener with the size metric being ineffective and the decay parameters showing the most potential.

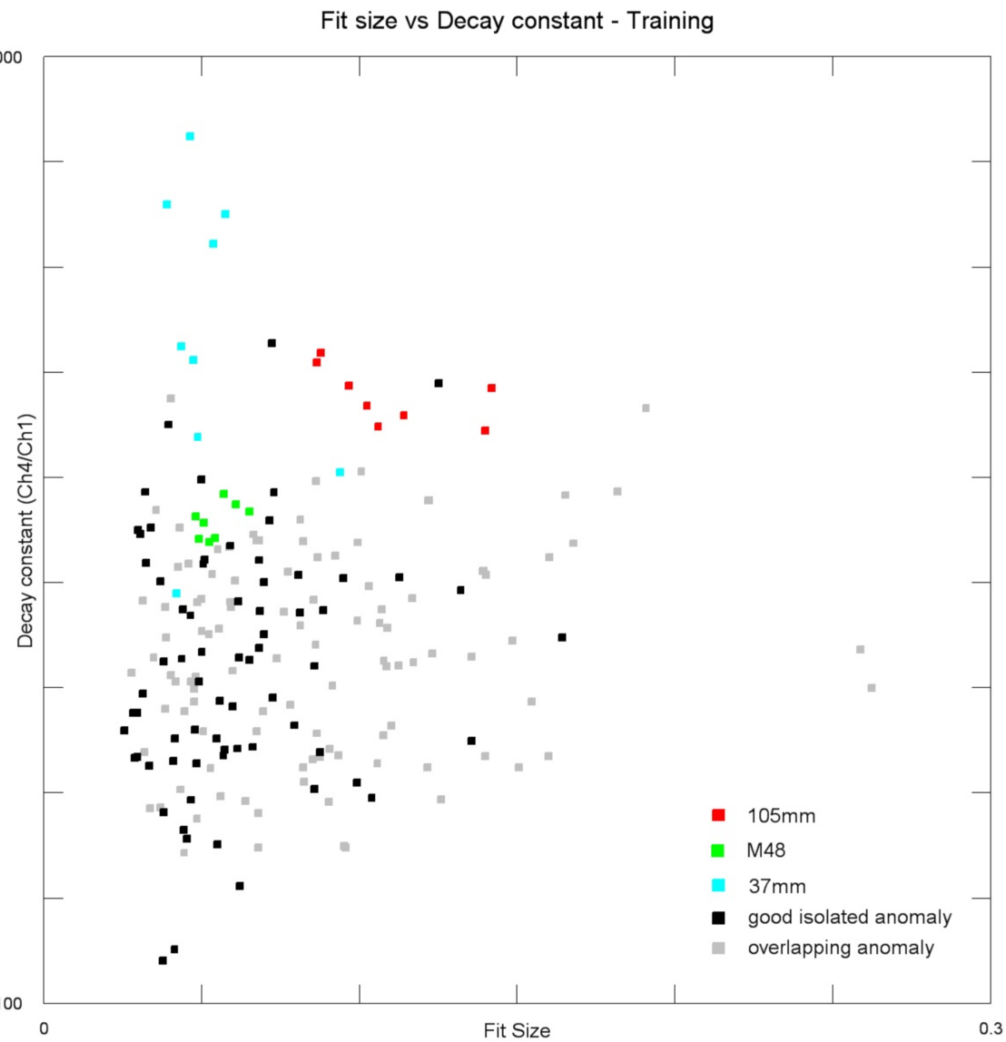


Figure 6-4. Inverted size parameter versus observed time decay for the standard training set and test pit measurements, EM61 cart data.

The final classification decision was a straight-forward, rules based, multi-step process based on thresholds calculated using the training data. Table 6-4 shows target parameters statistics for all the training TOI that had good data quality. We first assumed each anomaly was a Category 3 or dig. Next, for all targets with a data quality of "good" or "decent" we looked at the decay constant or TAU of each of the anomalies. If the number of points used to calculate TAU14 was greater than 4 we used it, if not we used TAU13 if it had greater than 4 points otherwise we classified the targets as Category 2 or "cannot decide". All anomalies with a TAU outside the mean ± 3*standard deviations of the TOI giving the highest or lowest TAU were classified as Category 1 or "clutter".

Each of the remaining anomalies with a fit error percentage of less than 45% and a data quality of "good" was compared to the TOI. If the anomaly had a TAU and fit size within the mean ± 3*standard deviations of a TOI it was classified as a Category 3 otherwise it was a Category 1.

All anomalies with a data quality other than "good" or "decent" were placed in Category 2. Finally, all anomalies that could not be inverted or had insufficient data coverage were placed in Category 4 or "cannot analyze".

Once all the anomalies were classified we ranked the anomalies within each of the categories and assigned a decision statistic based on how closed the anomaly's feature parameters were to the mean parameters that were used to classify the anomaly. For example if an anomaly was classified as a Category 3 anomaly matching a 105mm, the anomaly's parameters would be compared to the mean values of the TAU and fit size for a 105mm. Figure 6-5 shows the ROC curve of applying the classification procedures to the training data.

Table 6-4. EM61 target parameter statistics of all training data with a data quality rating of good.

	TAU14		TAU13		Fit Size	
TOI	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
37mm	790	101	548	54	.047	.006
M48 Fuze	678	32	414	43	.058	.013
105mm	672	28	508	16	.109	.021

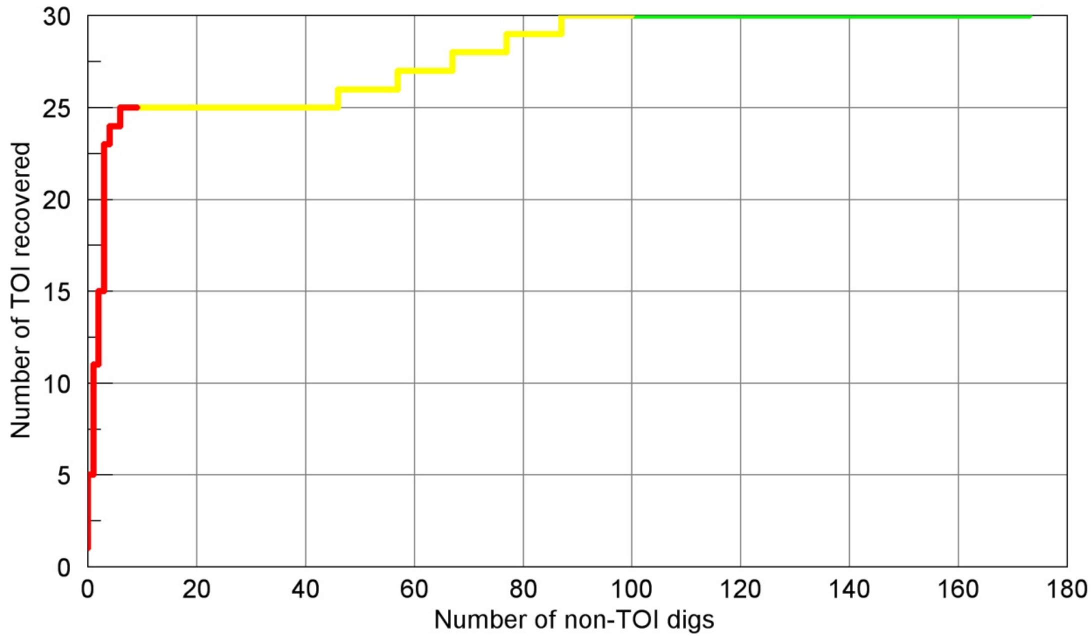


Figure 6-5. ROC curve of the EM61 cart training data. The plot is colored according to category with red, yellow and green corresponding to Category 3, 2 and 1, respectively.

6.4.2 TEMTADS Data

The TEMTADS data were analyzed using SAIC's custom IDL based software and UX-Analyze. Classification for both software packages was primarily based on an algorithm which compares our derived polarizabilities with a library of known target signatures. Some of the lists used all the polarizations to create 3 criteria: the amplitude of the primary polarizability (β_1), and two shape parameters calculated from the ratio of the second and third polarizability to the first (β_2/β_1 and β_3/β_1). Due to the occasional "splitting" of the two smaller β's, we developed the 2 criteria metric in which the 3rd comparison, namely the β_3/β_1 shape match, was dropped from the calculation. The difference in the values is computed at all time gates, excluding those where the values are negative. Finally, the results from the 2 or 3 different criteria were averaged, producing a metric which ranges from 0 (worst possible fit) to 1 (perfect fit). Note that the procedure just described is not a library constrained match, i.e., we do not invert our data forcing the β's to be those of each library object in turn, but rather simply compare our unconstrained polarizabilities to those of the library. As such, the comparison runs rapidly, and there is no need to reduce the number of separate types in the library to balance computation time. There were four IDL based submissions: (1a) A first pass run using no on-site training data and the 3-criteria beta matching, (1b) The second pass of the 1a which incorporated additional on-site training based on the results of 1a; (2) A dig list using training data and the 3-criteria beta matching; (3) A dig list using training data and the 2-criteria beta matching. The IDL based analysis used a single dipole model so care was taken to exclude Tx-Rx pairs that appeared to have interference from overlapping targets.

The UX-Analyze submittal using only the TEMTADS data used both the single- and multi-dipole models and kept the results that were a better match to a TOI. The submittal using the EM61 data as a pre-screener and TEMTADS data for the remainder of the anomalies used only the multi-dipole model. Both submittals used the 2-criteria beta matching.

Experience with this array has shown that the first six to eight timegates are affected by ringing from the transmitter. We therefore excluded these from consideration. In this section, we present scatter plots to provide a feel for the inverted target attributes.

Both our single- and multi-dipole models invert for the position, orientation, and principal axis polarizabilities (β's) as a function of time gate for the target(s). In Figure 6-6 to Figure 6-9, we show β plots for three TOI and a piece of clutter. Figure 6-6 shows a 105mm projectile. As expected for axially symmetric ordnance, there is one larger β and two smaller, equal ones. In Figure 6-7 and Figure 6-8, we plot a M48 fuze and 37mm projectile, respectively. These munitions also show good asymmetry at the early time gates but more variability later in time. In particular, the two smaller β's often show a greater separation as the measured data for these smaller ordnances gradually reach the sensor's noise levels. Finally, in Figure 6-9, we present the β plot for a sample clutter item. As is evident, the β's are quite distinct from those of the three ordnance items shown.

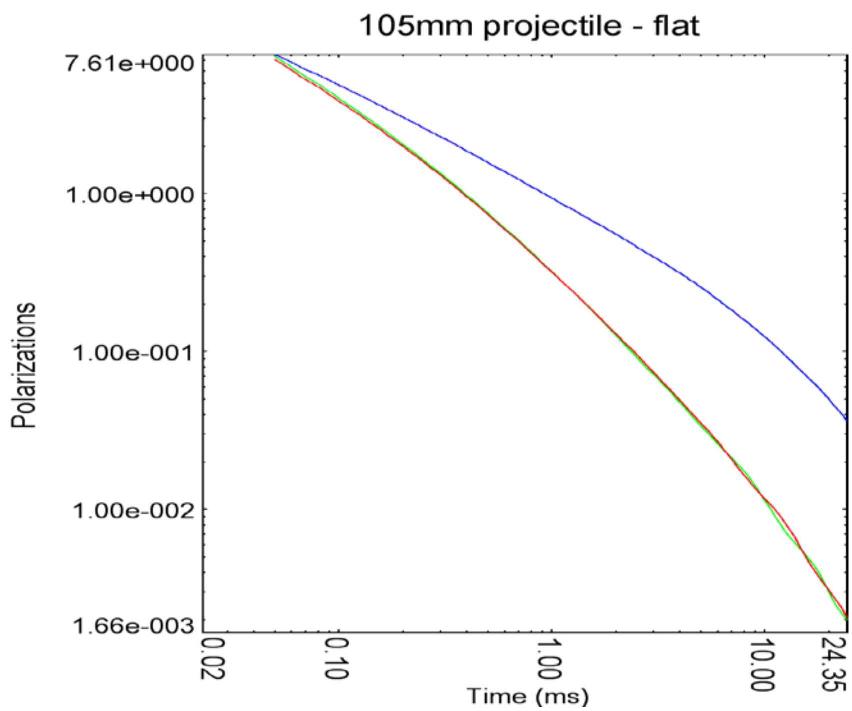


Figure 6-6. Principal axis polarizations (units of m^3) for a 105mm projectile (TEMTADS data).

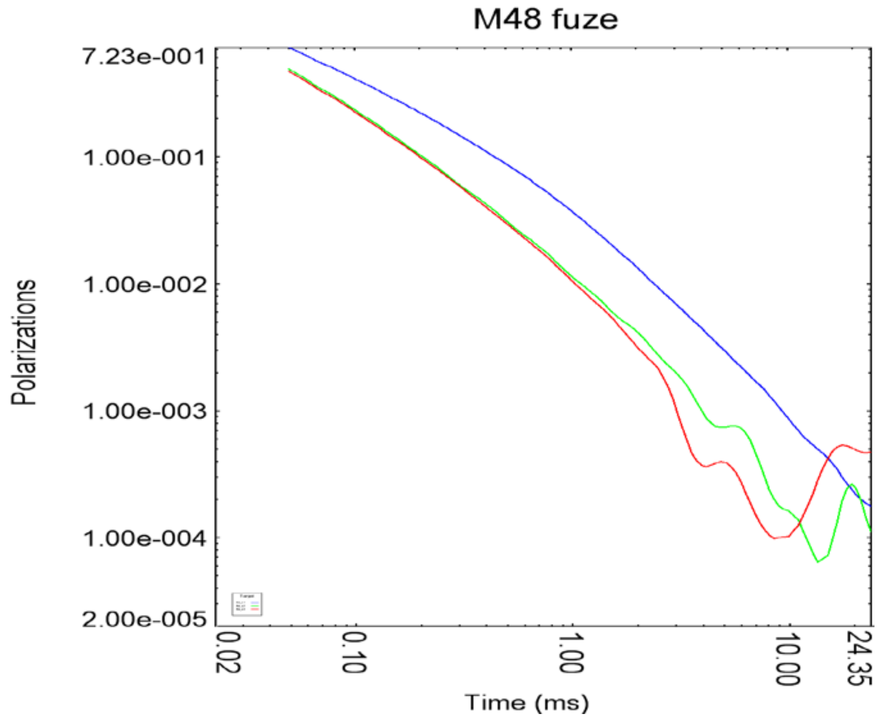


Figure 6-7. Principal axis polarizations (m^3) for a M48 fuze using TEMTADS data.

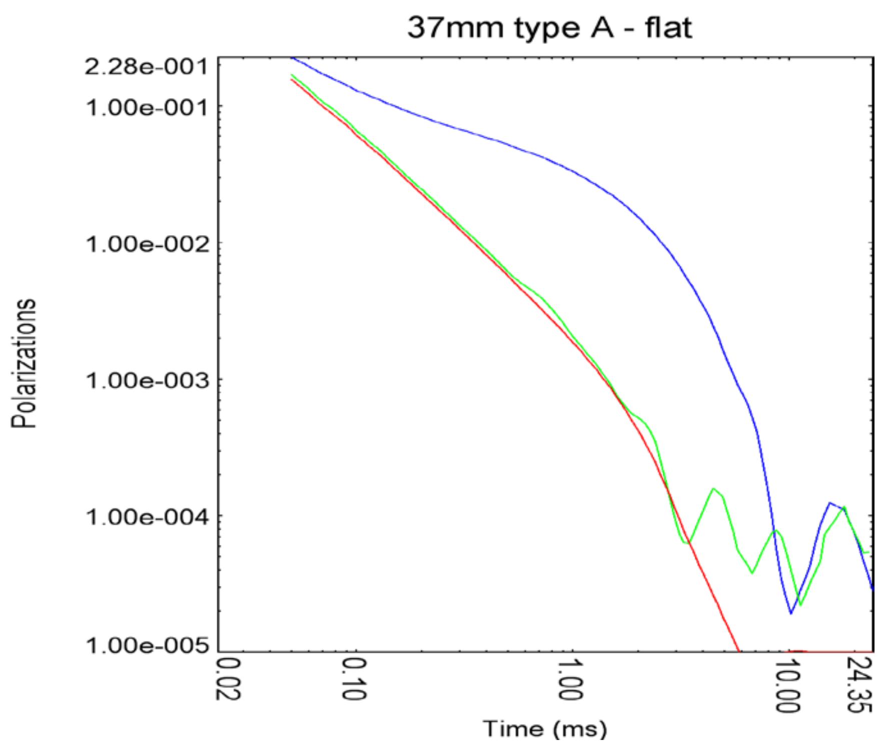


Figure 6-8. Principal axis polarizations (m^3) for a 37mm projectile using TEMTADS data.

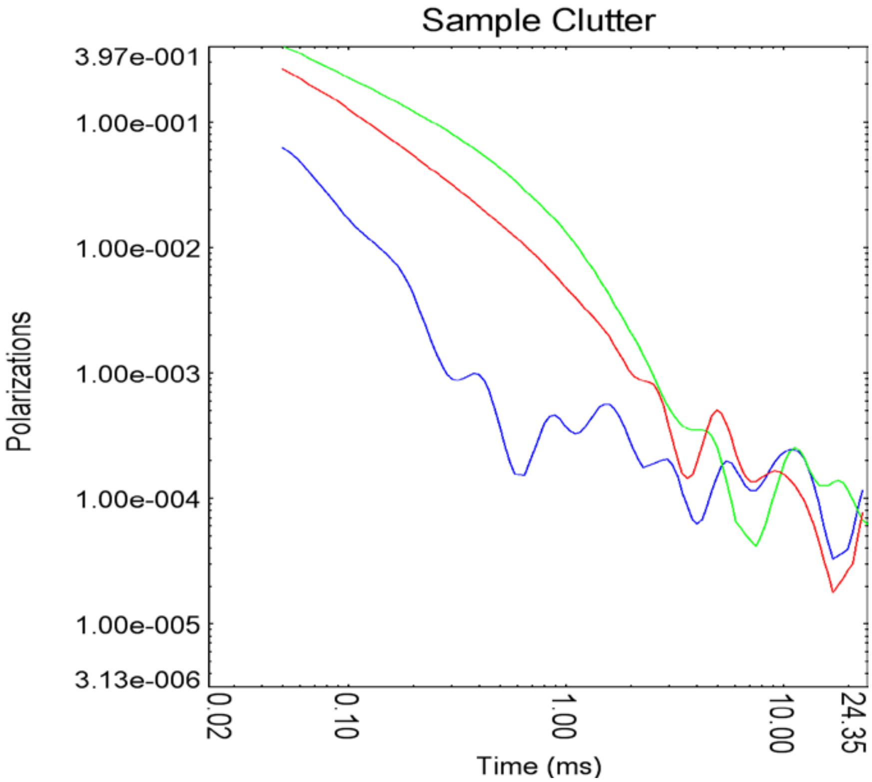


Figure 6-9. Principal axis polarizations (m^3) for a clutter object using TEMTADS data.

In the majority of cases, these library signatures were taken under test stand conditions. The library only contained samples of the three ordnance types known to be present at the site. It consisted of measurements obtained at various target orientations, as well as several subtypes of these munitions. It also included measurements from the site test pit whenever the measured signature appeared distinct from those of the same type already present.

The different classification methods use various amounts of training data. Two IDL based analysis methods (Table 6-1, dig list 2 and 3) used the same set of training data. The requested targets was based on 4 separate categories: (1) Anomalies for which the derived polarizabilities produced a good match to an ordnance target in our full library, where the target is not one of the expected ordnance items at the site; (2) Anomalies whose polarizabilities looked “UXO-like”, although they didn’t match any target in our library well; (3) Anomalies for which the library match metric fell within a range slightly below the standard range for good library matches and ; (4) Anomalies which fall in the middle range of our metric scale, and for which the match to our UXO library was comparable to the match to a library based on clutter measurements. We requested 73 anomalies of which 29 were UXO.

The two NOSLN approaches (Table 6-1, dig list 1 and 9) did not request any training data. Their thresholds were determined by the poorest match to the ordnance types known to be present,

based on previous measurements and experience. A visual inspection of matches in the Test Set was also conducted to adjust the threshold. An additional submittal was generated which was essentially the 2nd pass for the IDL NOSLN approach. For this submittal we requested training data based on the classification results for the 1st pass. All targets classified as "Likely Munitions" from the first pass were requested, along with roughly 10% each of targets classified "Cannot Decide" due to falling in the "buffer zone", and due to possessing axial symmetry. This resulted in 168 anomalies of which 117 were UXO. The additional training anomalies did not uncover any unexpected results therefore our thresholds did not change for the second pass.

In this subsection, we describe our decision rules for classifying anomalies into the four categories in the final dig list, as well as our criteria for determining overlapping signatures. All the different TEMTADS analysis approaches used very similar decision rules to categorize the anomalies but slightly different thresholds. The rules for the IDL 2 criteria and 3 criteria approaches are detailed below.

We classified as "Cannot Analyze" targets for which the inversion produced unphysical parameters, specifically, depths below a 2m cutoff which was based on the largest target expected on the site, and negative polarizabilities.

We classified targets as "Likely Munitions" based on the library match metrics previously described. The cutoff was determined to insure that all training targets revealed as TOI with moderate-to-good signal-to-noise were included. Our cutoff values were 0.6944 for the 3 criteria metric, and 0.7849 for the 2 criteria metric. For the NOSLN approaches our cutoffs were also altered based on visual inspection of the library matches for the test anomalies. Table 6-5 shows the different thresholds used for each of the analysis approaches.

Targets in the "Cannot Decide" category were comprised of three distinct types. The first type consisted of targets with very low signal-to-noise. Based on previous experiences with this instrument a value of 1.5mV/A was used. The second type was targets which suffer from serious overlap issues with neighboring anomalies, to the extent that we were not confident in our ability to extract meaningful features. The remaining category excludes cases from the first two. It consists of anomalies for which the metric falls below the "Likely Munitions" cutoff, but which possess axial symmetry. This choice is made to take account of the fact that our library is finite. Exceptions to the axial symmetry rule were targets showing rapid decay (based on the most rapid decay for UXO in the complete library), and targets whose polarizabilities were of very small amplitude (based on the smallest amplitude UXO in the complete library). We conservatively defined axial symmetry as a median agreement of 50% between the two smaller betas.

We placed targets in the "Likely Clutter" category whose metric fell below the "Likely Munitions" cutoff and which do not show low signal-to-noise, serious overlap, or possess axial symmetry, as previously defined.

The "Likely Clutter" category was divided into two: low signal-to-noise and moderate to high signal-to-noise. Within each of these subcategories, ranking were by the library match metric, such that the lower the metric, the higher the rank. However, all anomalies in the low SNR

subcategory were ranked lower than those in the moderate to high SNR subcategory. The rationale behind this was that low SNR can decrease the quality of the match metric, and the targets for which we have the highest confidence of being clutter are those with a low metric and decent SNR. Similarly, the "Cannot Decide" category was divided into three based on the types of entries, with the targets having low signal and overlap issues having the lowest rank, followed by those with moderate-to-high signal and no serious overlap issues, ending with those targets having a metric below the "Likely Munitions" cutoff, but which possess axial symmetry. The ranking within each subcategory were by the metric. Finally, the "Likely Munition" category was ranked solely by the metric, with the largest rank going to the largest metric value.

Since it was not obvious how to incorporate either the axial symmetry parameter or the low SNR parameter into our probabilities, we set the probability equal to 1 minus the metric value for all targets. Unfortunately, with the exception of the "Likely Munition" category, this means that there was not in general a 1 to 1 mapping between the ranks and the probabilities.

Contour plots were made of each anomaly both as a QC check and as a means of determining whether the anomaly suffers from overlapping signatures. An entry was made in a spreadsheet giving some indication of the extent of the overlap, if any. However, in some cases, this overlap may be from multiple targets within the anomaly, as opposed to being caused by a separate anomaly. Therefore, we used this entry in conjunction with the knowledge of the location of other anomalies relative to the target of interest. The edge of the TEMTADS is 1m from a target located under the array center. Thus, we declared any anomaly for which the spreadsheet entry indicates overlap, and for which at least one target is within 1.5m of the anomaly, as an overlapping signal. The UX-Analyze analyses used the multi-dipole solver therefore did not need to perform this check. Instead the inverted results for all the objects found at particular anomaly flag were analyzed and the object with the best match to a TOI was kept.

In order to match a recovered object to the respective flag, there was requirement that the excavated object must be within 60cm of the flagged location. The TEMTADS array being 2m square can detect other anomalies while collecting data for a specific flag. Typically all the measured data is passed to the inversion algorithms and in the case of the multi-dipole solver the locations of several objects may be returned. For the UX-Analyze analyses we made a requirement that only objects with a fitted location less than 80cm from the flagged location were considered in our classification of the flagged anomaly. Our experience has also shown that if the fitted location were along the edges of the array the fitted results are not as reliable. Therefore we made an additional requirement that the fitted location must also be within 0.65m of the center of the array in order to make a classification decision based on the fitted results.

The two NOSLN approaches used the same procedures described above but also included an additional subcategory to the "Cannot Decide" category. It consisted of targets whose match metric fell within a "buffer zone" below the "Likely Munitions" metric cutoff. The anomalies in this category possess a match metric which is not sufficiently high to justify their being classified as "Likely Munitions", but which is not sufficiently low to classify them as "Likely Clutter".

Table 6-5. Thresholds used for the different analysis approaches using TEMTADS data.

	Metric Threshold	Buffer Threshold	Axial Symmetry Metric	Distance to Flag Metric	Distance from Array Center Metric
IDL 2crit	0.7849	N/A	50%	N/A	0.8
IDL 3crit	0.6944	N/A	50%	N/A	0.8
IDL NOSLN	0.8465	0.6517	50%	N/A	0.8
IDL NOSLN	0.8465	0.6517	50%	0.8	0.8
UXA NOSLN	0.89	0.81	75%	0.8	0.65

6.4.3 Metal Mapper Data

The Metal Mapper data were analyzed using SAIC's custom IDL based software and UX-Analyze. Similar to the TEMTADS analysis, classification was primarily based on an algorithm which compares our derived polarizabilities with a library of known target signatures as described in the previous section. There were four IDL based submissions: (1a) A first pass run using no on-site training data and the 3-criteria beta matching, (1b) The second pass of the 1a which incorporated additional on-site training based on the results of 1a; (2) A dig list using training data and the 3-criteria beta matching; (3) A dig list using training data and the 2-criteria beta matching. The IDL based analysis used a single dipole model.

The UX-Analyze submittal used the EM61 data as a pre-screener and Metal Mapper data for the remainder of the anomalies. It used both the single- and multi-dipole models and kept the results that were a better match to a TOI. Classification was based on the 2-criteria beta matching. As with the TEMTADS, 319 anomalies were classified with the EM61 data and the remaining 1971 were based on the MM data.

In Figure 6-10 to Figure 6-13, we show β plots for three TOI and a piece of clutter. Figure 6-10 shows a 105mm projectile. As expected for axially symmetric ordnance, there is one larger β and two smaller, equal ones. In Figure 6-11 and Figure 6-12, we plot a M48 fuze and 37mm projectile, respectively. Finally, in Figure 6-13, we present the β plot for sample clutter item. Similar to the TEMTADS data the TOI shows good axial symmetry and the clutter object produced β's that are quite distinct from those of the three ordnance items shown.

Classification for the Metal Mapper dataset used a library of known target signatures obtained with the TEMTADS. Both the Metal Mapper and the TEMTADS were developed at G & G Sciences, and have essentially the same waveform. Experience with the precursor of the Metal Mapper, the AOL2, demonstrated that the polarizabilities derived from that system could be well-matched to ones derived from TEMTADS data by the application of a time lag and scale factor to the former. We used Metal Mapper data taken in the test pit to determine the appropriate lag and scaling correction to match the TEMTADS library to the Metal Mapper data. For each target in the TEMTADS library we followed the following steps prior to the library matching: First, the time lag derived as described above was applied to the TEMTADS timegates. Second, the derived scale factor was applied to the TEMTADS polarizabilities. Third, any TEMTADS timegates occurring later or earlier than those of the Metal Mapper were discarded. Finally, the TEMTADS polarizabilities were interpolated to the Metal Mapper timegates.

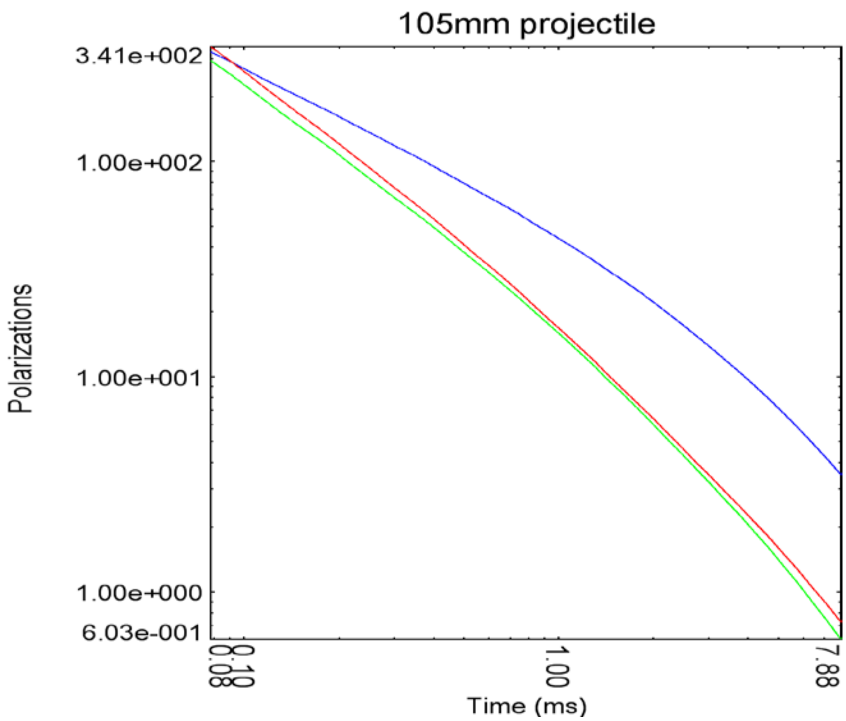


Figure 6-10. Principal axis polarizations for a 105mm projectile using Metal Mapper data.

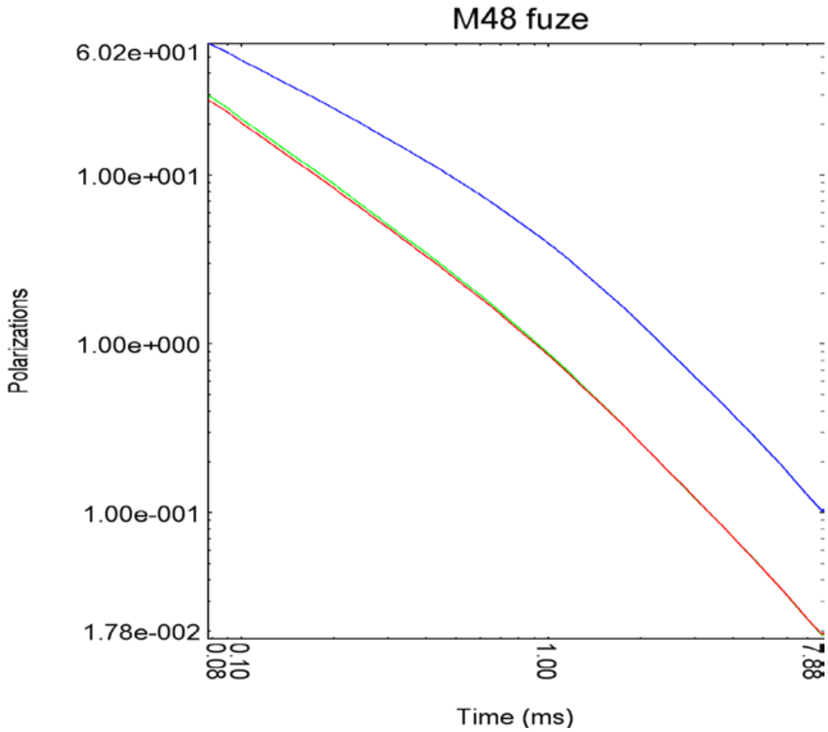


Figure 6-11. Principal axis polarizations for a M48 fuze using Metal Mapper data.

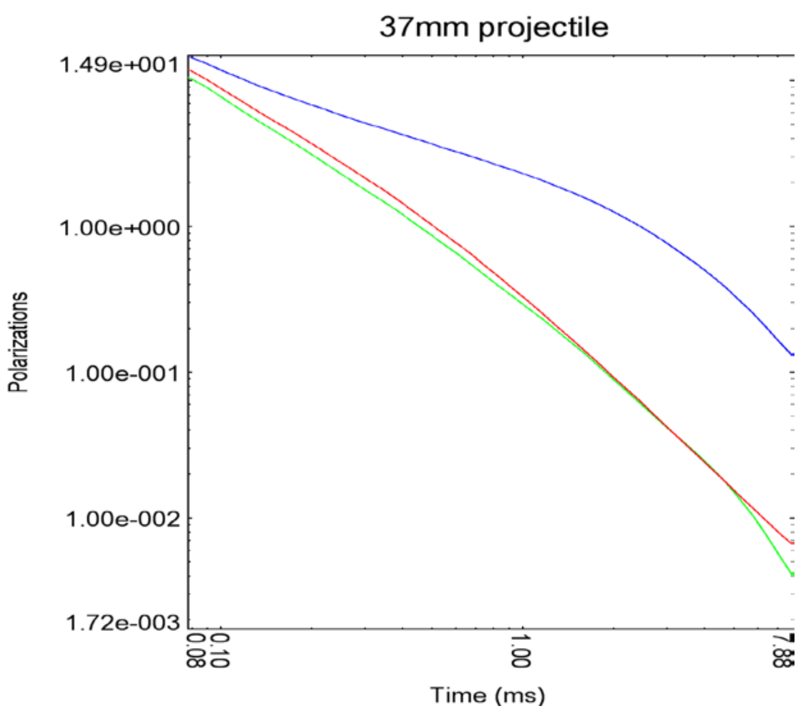


Figure 6-12. Principal axis polarizations for a 37mm projectile using Metal Mapper data.

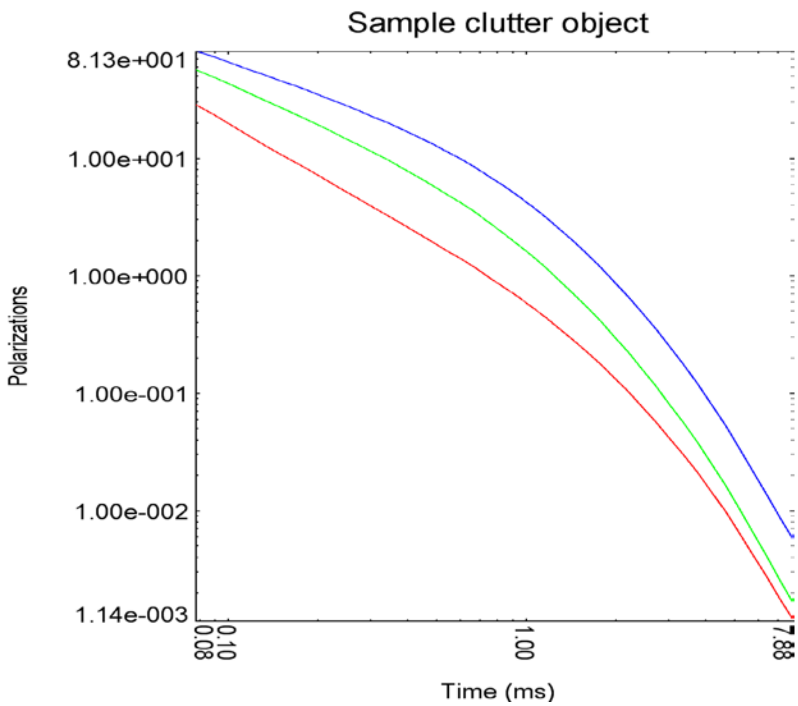


Figure 6-13. Principal axis polarizations for a sample clutter object using Metal Mapper data.

In the majority of cases, the signatures in the library were taken under test stand conditions. The library contains a number of samples of the three ordnance types known to be present at the site, including measurements obtained at various target orientations, as well as several subtypes of these munitions. The library items were a combination of items measured at previous sites and measurements taken from the test pit at this site. Whenever a measured signature from the test pit appeared distinct from those of the same type already present, we added it as a separate entry.

The different classification methods use various amounts of training data that were chosen in the same fashion as the TEMTADS analysis described in the previous section. The 2-criteria methods used 110 anomalies for training of which 10 were UXO. The 3-criteria method requested 112 anomalies for training with 9 being UXO.

The two NOSLN approaches did not request any training data. Their thresholds were determined by the poorest match to the ordnance types known to be present, based on previous measurements and experience. A visual inspection of matches in the Test Set was also conducted to adjust the threshold. An additional submittal was generated which was essentially the 2nd pass for the IDL NOSLN approach. For this submittal we requested training data based on the classification results for the 1st pass. All targets classified as “Likely Munitions” from the first pass were requested, along with roughly 10% each of targets classified “Cannot Decide” due to falling in the “buffer zone”, and due to possessing axial symmetry. This resulted in 296 anomalies of which 149 were UXO. Using the additional training anomalies slightly lowered the “Dig”/”No Dig” threshold for the second pass.

The decision rules for classifying the Metal Mapper anomalies were the same as those described for the TEMTADS except different thresholds were used. For the "Likely Munitions" category the cutoff values were 0.65 for the 3-criteria metric, and 0.71 for the 2-criteria metric. Table 6-6 shows the different thresholds used for each of the analysis approaches. The "Cannot Decide" category included targets that; a) suffered from serious overlap issues with neighboring anomalies, b) possess axial symmetry, and c) have low signal-to-noise. The "Likely Clutter" category consisted of targets whose metric fell below the "Likely Munitions" cutoff and which do not show low signal-to-noise, serious overlap, or possess axial symmetry, as previously defined.

The targets were ranked in a similar fashion to the TEMTADS ranking. Targets within each of the three "Can Analyze" categories were ranked solely by their metric value, with the largest rank going to the largest metric value. The "Cannot Decide" category was divided into three based on the types of entries as described above. Targets having overlap issues and low SNR were given the lowest rank, followed by targets possessing axial symmetry. Again, the ranking within each of the two subgroups were based solely on the metric.

Although the Metal Mapper receiver cubes are spread out in various locations, the design of the system does not allow any obvious means of obtaining a contour plot for a given anomaly analogous to systems with single-plane transmitters and receivers. Thus there is no visual clue as to whether overlap occurs. Therefore, we simply used the knowledge of the location of other anomalies relative to the target of interest. The furthest receiver cubes are only 40cm from the center of the sensor, well within the 60cm used to declare whether anomalies should be considered separate or not. Thus, we declared any anomaly for which at least one target was within 1.5m of the anomaly, as an overlapping signal.

The UX-Analyze NOSLN approach used the same procedures described above but also included an additional subcategory to the "Cannot Decide" category. It consisted of targets whose match metric fell within a "buffer zone" below the "Likely Munitions" metric cutoff. The anomalies in this category possess a match metric which is not sufficiently high to justify their being classified as "Likely Munitions", but which is not sufficiently low to classify them as "Likely Clutter".

Table 6-6. Thresholds used for the different analysis approaches using MM data.

	Metric Threshold	Buffer Threshold	Axial Symmetry Metric	Distance to Flag Metric	Distance from Array Center Metric
IDL 2crit	0.71	N/A	50%	N/A	N/A
IDL 3crit	0.65	N/A	50%	N/A	N/A
IDL NOSLN 1st pass	0.69	N/A	50%	N/A	N/A
IDL NOSLN 2nd pass	0.65	N/A	50%	N/A	N/A
UXA NOSLN	0.89	0.81	75%	0.8	0.65

6.5 PERFORMANCE ASSESSMENT

All submitted dig lists were scored against the emplaced and recovered targets by the IDA. An example ROC curve is shown in Figure 6-14, with the areas of interest for the analysis indicated.

The scoring software plotted the Percent of Munitions Dug versus the Number of Unnecessary Digs for all possible dig thresholds and drew vertical grey bars around each point on the ROC curve to denote the 95% confidence interval around the point's Percent of Munitions Dug value. The scoring software colored in black the extra point on the ROC curve corresponding to the "Dig Everything" situation, in which ALL locations in the Test Set are declared "Dig"; this point is always at the upper, right end of the ROC curve. Conversely, the software also colored in black the extra point corresponding to the "Leave Everything in the Ground" situation, in which ALL locations in the Test Set are declared "Do not dig", including the Cannot Analyze locations. This point is always at the lower, left end of the ROC curve. Some of the curves have a gap between the origin and the black dot at the lower left end of the curve. This represents the training set anomalies. From a declared-category perspective, the Category 4 targets are plotted first (black), followed by categories 3 (red), 2 (yellow), and finally 1 (green). The colored dots on the ROC curves indicate the operating point for a Pd=0.95 (pink), the demonstrator's threshold point (dark blue) and the "best case scenario" dig threshold which has the lowest number of FP for Pd=1.0 (light blue).

In a real-world situation, all locations in the Training Set must be dug. Training set locations that were truly of the munition type of interest cause a constant shift increase in the Percent of Munitions Dug values plotted along the Y axis. Similarly, training set locations that were truly clutter cause a constant shift increase in the Number of Unnecessary Digs values plotted along

the X axis. Including the training set locations allows for an apples-to-apples comparison between ranked dig lists created from training sets of different sizes, as all ROC curves based on the same data collection instrument will have their upper, right "Dig Everything" point in the same location, regardless of the size of the training set used.

ROC curves were generated for each anomaly list submitted. Each data set will be discussed individually in the sections that follow. As part of the discussion a number of figures are presented comparing the inverted parameters to ground truth information for all excavated anomalies. Due to the large number of anomalies, we have segmented the plots based on our classification categories. This also serves to illustrate differences between the categories. In all figures, UXO/TOI are plotted in red, while clutter are plotted in black. In general the agreement with ground truth improves with increasing category number. This is presumed to be due to a large number of small, low SNR frag items in Category 1, which result in greater uncertainty in both the measured and fitted values. Also, UXO objects being compact in size and shape tend to fit our point dipole model better than some clutter objects which can be quite irregular in shape.

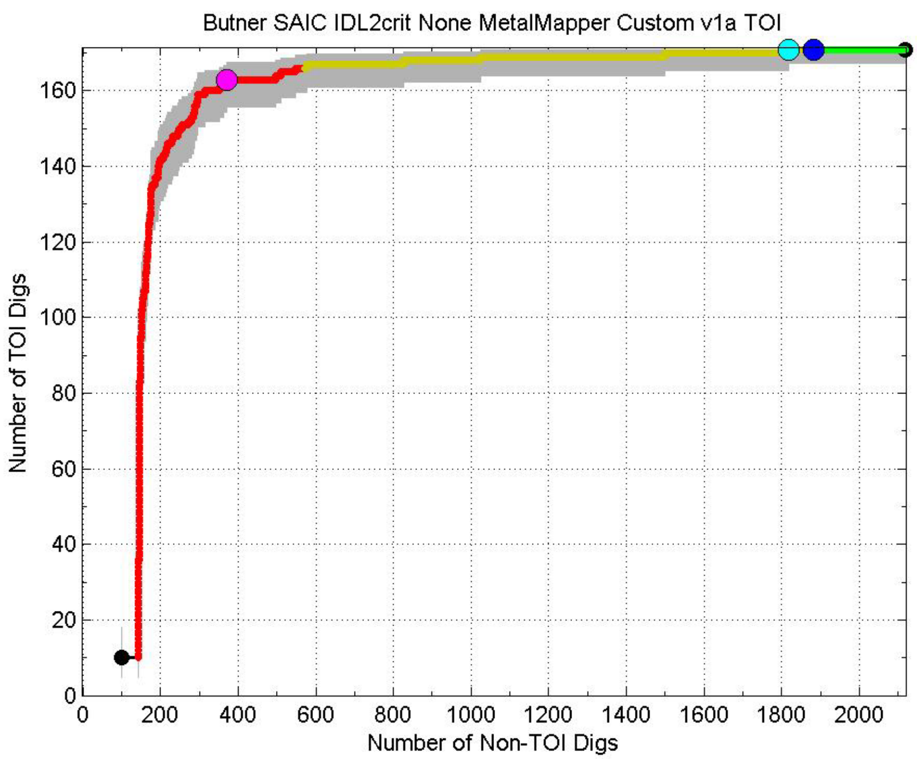


Figure 6-14. Example ROC curve with black, red, yellow and green lines representing category 4, 3, 2, and 1 anomalies, respectively. The colored dots on the ROC curves indicate the operating point for a Pd=0.95 (pink), the demonstrator's threshold point (dark blue), the "best case scenario" dig threshold which has the lowest number of FP for Pd=1.0 (light blue) and the amount of training anomalies used (black).

6.5.1 UX-ANALYZE - EM61 MK2 CART

Data from the EM61 MK2 cart is shown in Figure 6-15. The crosses identify anomalies selected for Test Set by the ESTCP Program Office.

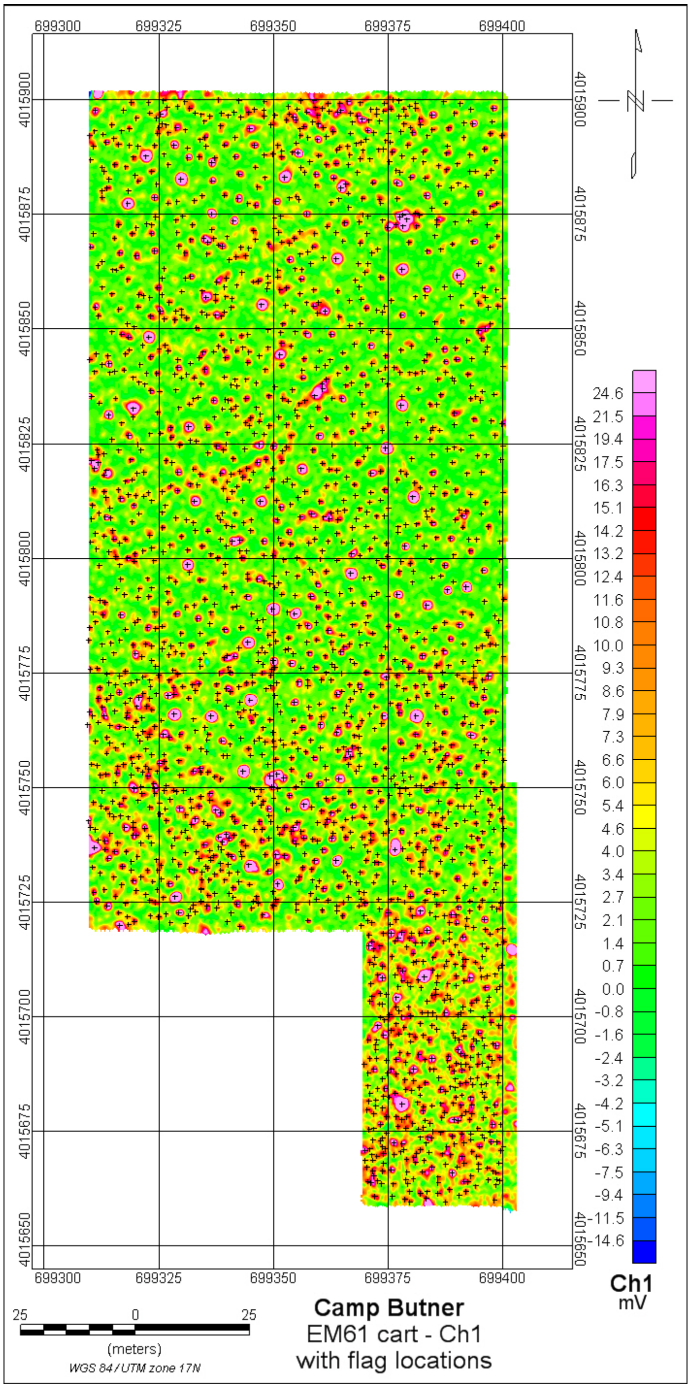


Figure 6-15. EM61 MK2 Cart data with selected test anomalies.

Performance Scores from IDA

Scoring performances for the EM61 MK2 cart analysis are reported in Table 6-7 A ROC chart is shown in Figure 6-16, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there was one false negative. Anomaly #404, which was a seeded 37mm, was classified as high confidence clutter (Figure 6-17).

Table 6-7 Test Set Summary: EM61 MK2 Cart

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	19	293	8	1	**321**
2	14	1073	25	18	**1130**
3	10	497	0	146	**653**
4	0	9	0	0	**9**
Training	0	164	7	6	**177**
TOTAL	**43**	**2036**	**40**	**171**	**2290**

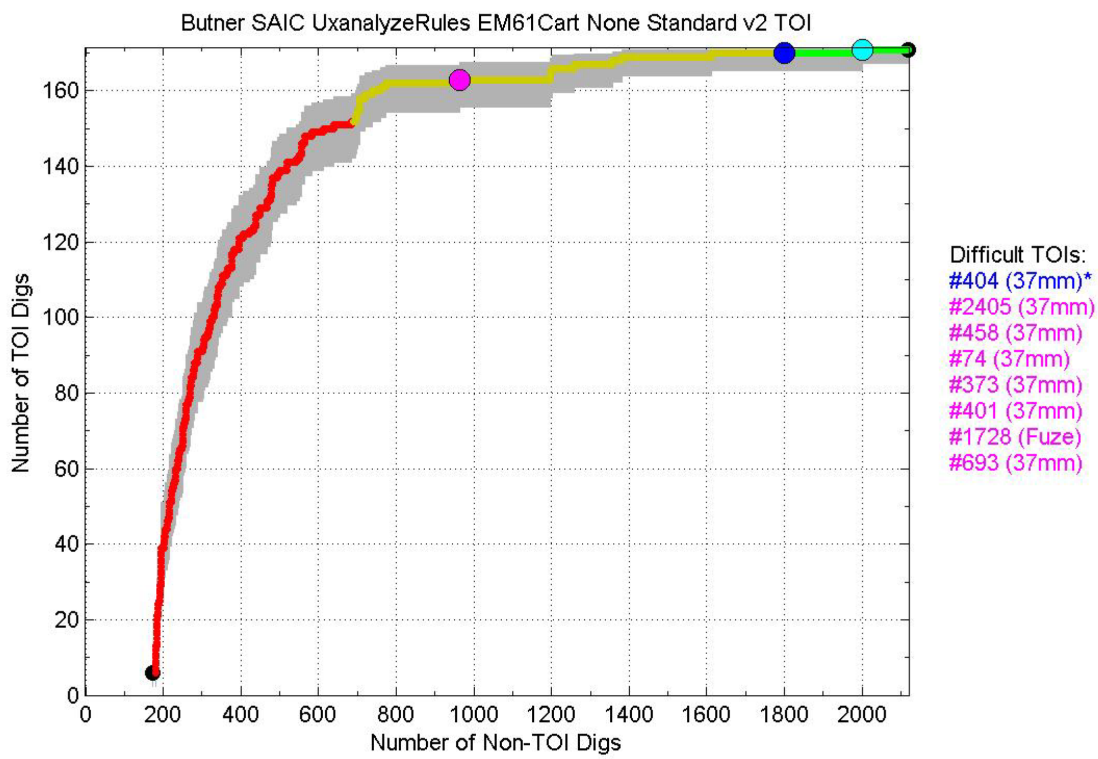


Figure 6-16. SAIC's EM61 MK2 Cart ROC chart.

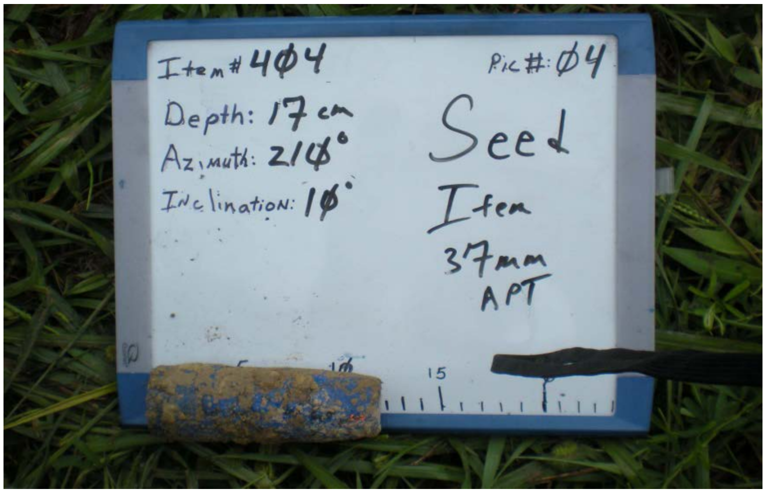


Figure 6-17. Photograph of target 404 which is a 37mm that was classified as high confidence clutter.

Characterization Plots

Figure 6-18 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.17m with a standard deviation of 0.12m. If non-TOI are added to the population the mean error increases to 0.18m with a standard deviation of 0.12m.

Figure 6-19 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.15m with a standard deviation of 0.14m. If non-TOI are added to the population the mean error increases to -0.18m with a standard deviation of 0.12m.

Figure 6-20 and Table 6-8 show the inverted polarizabilities for all category 1, 2 and 3 targets from the Test Set. In general, the calculate size (sum of betas) of the main TOI shows much smaller relative deviations than the individual betas.

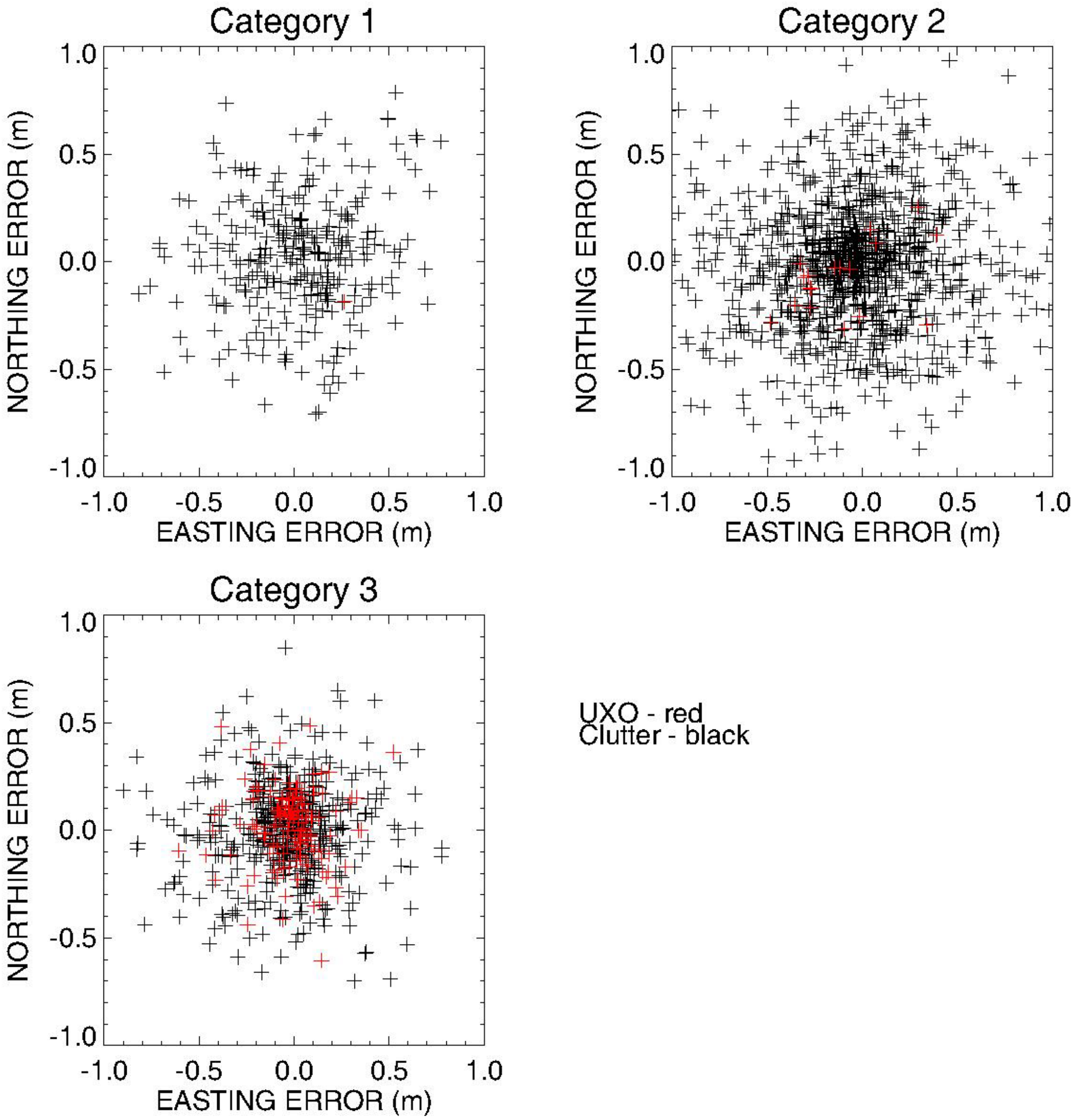


Figure 6-18. Differences between fitted and measured XY locations; EM61 MK2 cart

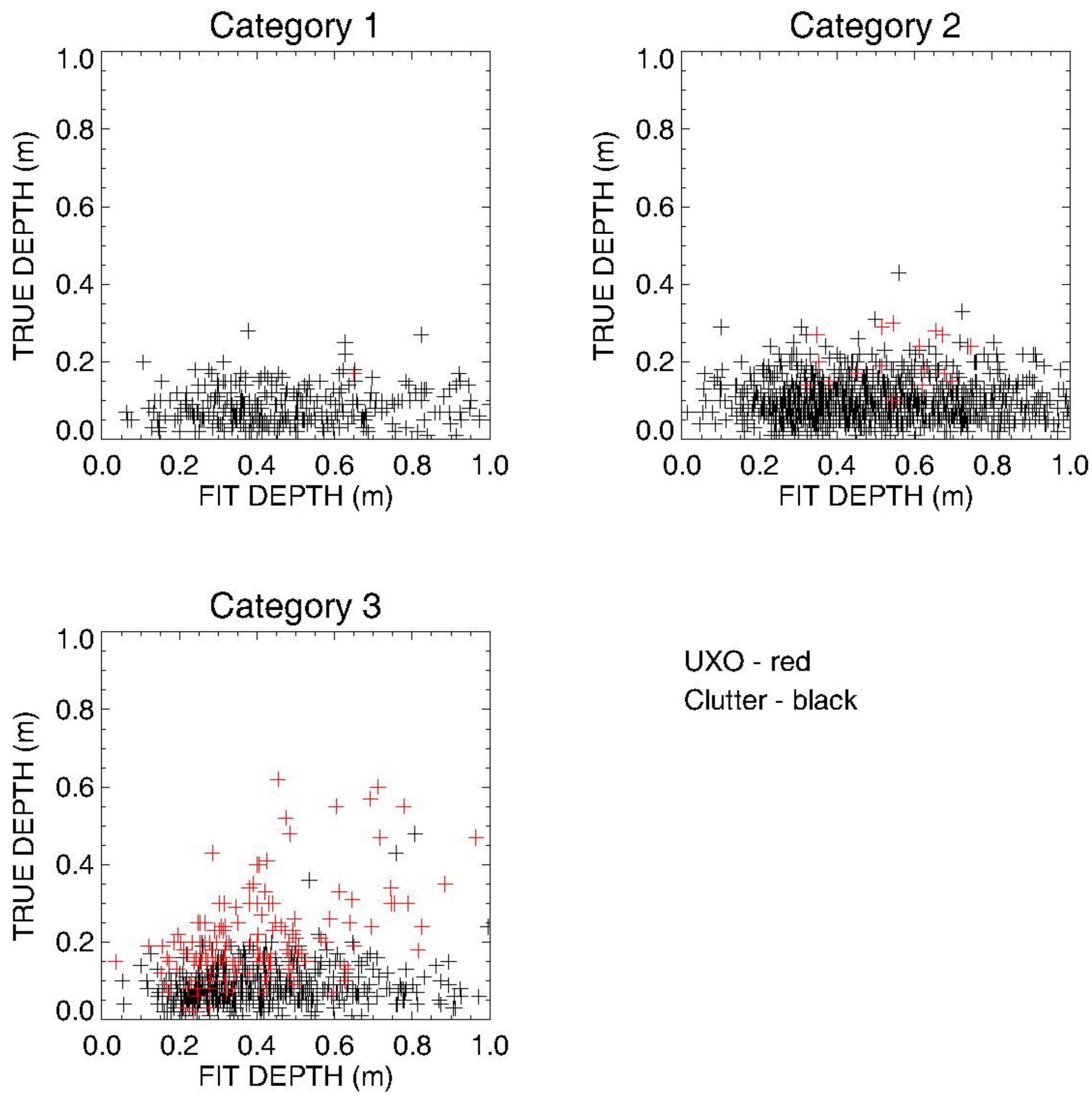


Figure 6-19. Fitted versus measured depth of burial; EM61 MK2 cart

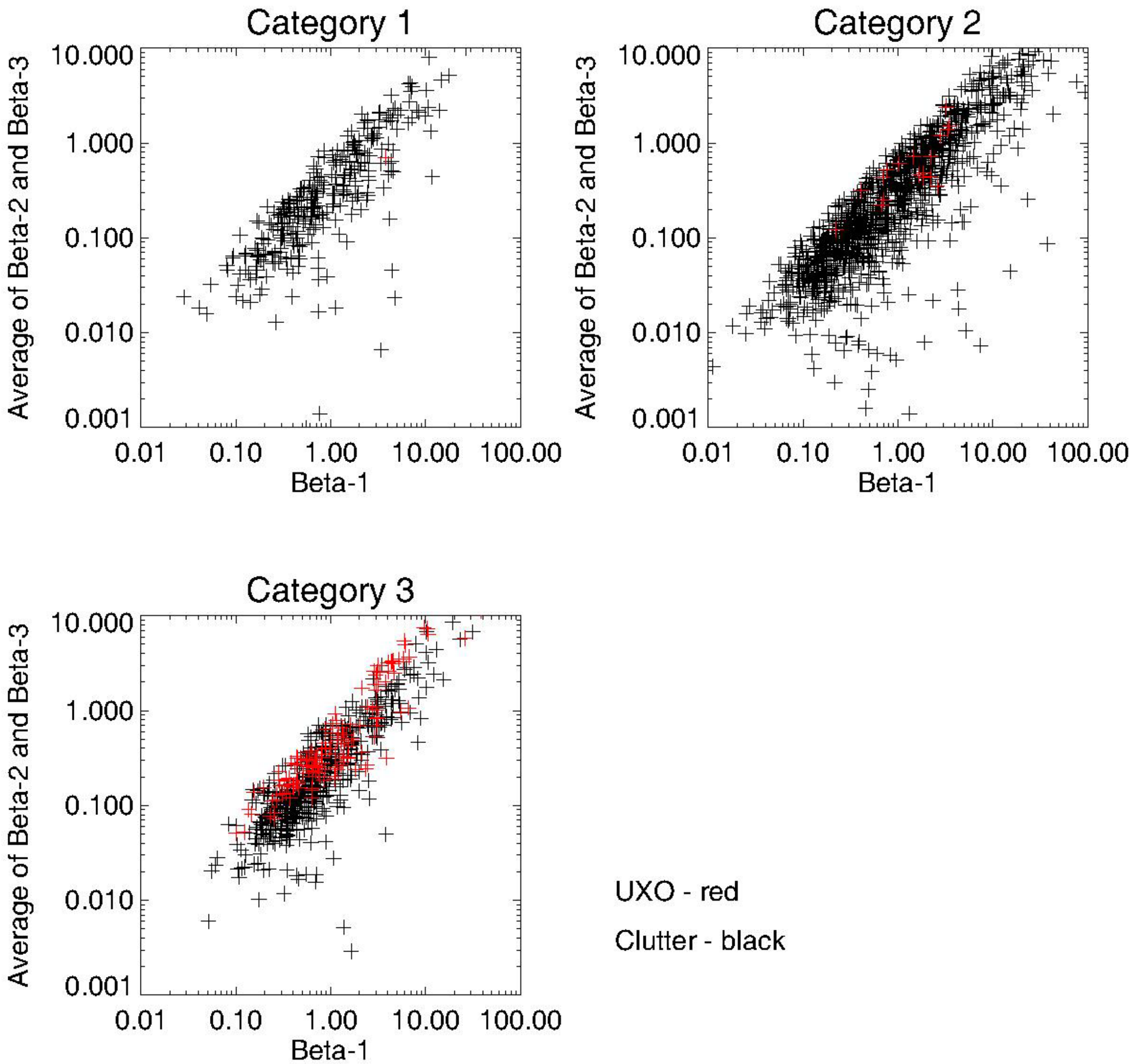


Figure 6-20. Beta 1 versus the average of Beta 2 and Beta 3; EM61 MK2 cart

Table 6-8 Statistics of Betas for the three main TOI, EM61 MK2 cart

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
37mm	113	0.055	0.014	1.120	1.076	0.397	0.359	0.111	0.100
Fuze	22	0.061	0.015	1.518	1.431	0.477	0.364	0.178	0.230
105mm	26	0.115	0.024	7.124	7.925	3.935	2.190	2.645	1.136

Failure Analyses

Anomaly #404 was a seeded 37mm but was classified as high confidence clutter (Category 1). A reexamination of the anomaly data revealed that both the TAU14 and TAU13 decay parameters indicated a high confidence non-TOI. The TAU14 parameter was 404 which was smaller than the 487 cutoff but had a standard deviation of 107. Figure 6-21 shows the measured and modeled data for anomaly 404. On careful examination, there we see the TAU14 value was around 575 over the center of the anomaly but was much smaller and below the cutoff in the north and eastern portions. This indicates that there may be several small fast decaying sources surrounding the anomaly. The high standard deviation should have prevented this anomaly from being declared high confidence clutter but a bug in the source code let this anomaly through.

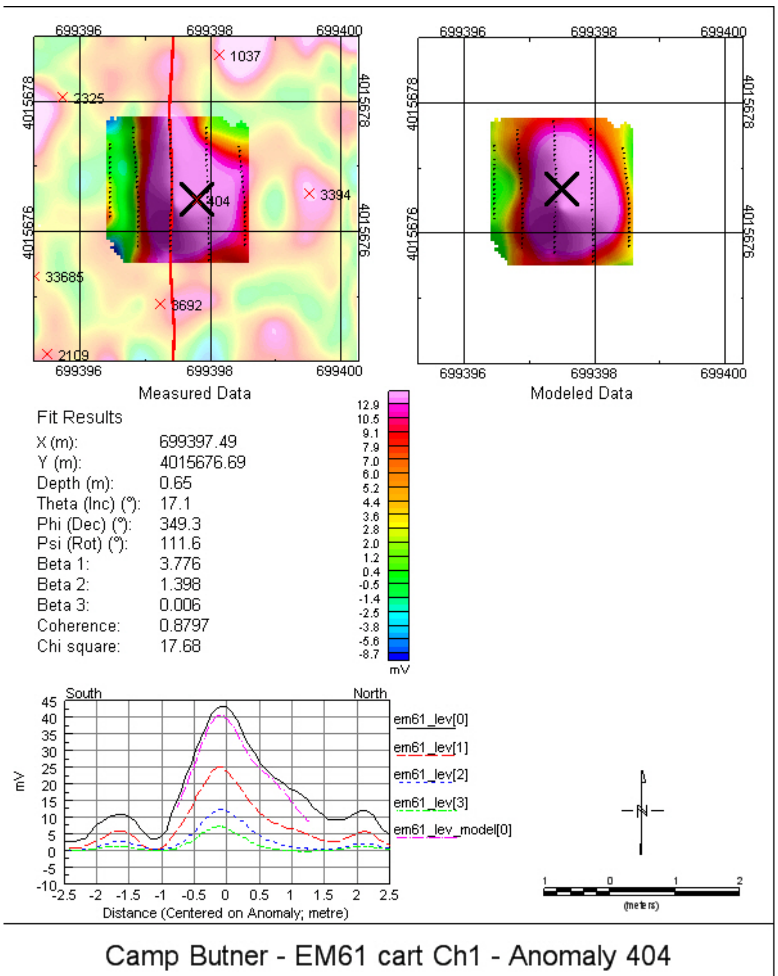


Figure 6-21. Anomaly plot showing measured data, inverted features and forward model for anomaly 404 which was incorrectly ranked as high confidence clutter.

6.5.2 UX-ANALYZE - EM61 MK2 CART - PARSONS

Parsons conducted a number of different analysis methods using the EM61-MK2 cart data. They used UX-Analyze to invert the data using a dipole model to extract target features. They also calculated decay parameter using UX-Analyze and an algorithm developed by Parsons. Specific details on the different processing methods and failure analysis can be found in Parsons' report which is included in this report as Appendix C.

Performance Scores from IDA

A ROC chart showing scoring performance for the EM61 MK2 cart analysis, which performed the best according to Parsons' final report, is displayed in Figure 6-22, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs. This list was deemed best based on IDA's retrospective 95% and 100% Probability of Detection "Don't Dig" thresholds.

Using the thresholds adopted for this analysis, there were 44 false negative. The high number of false negatives was due to an aggressive choice of the "Dig" / "Don't Dig" threshold. Retrospectively, if a more conservative threshold were picked a Probability of Detection greater than 95% could be achieved while still leaving over 40% of the clutter in the ground.

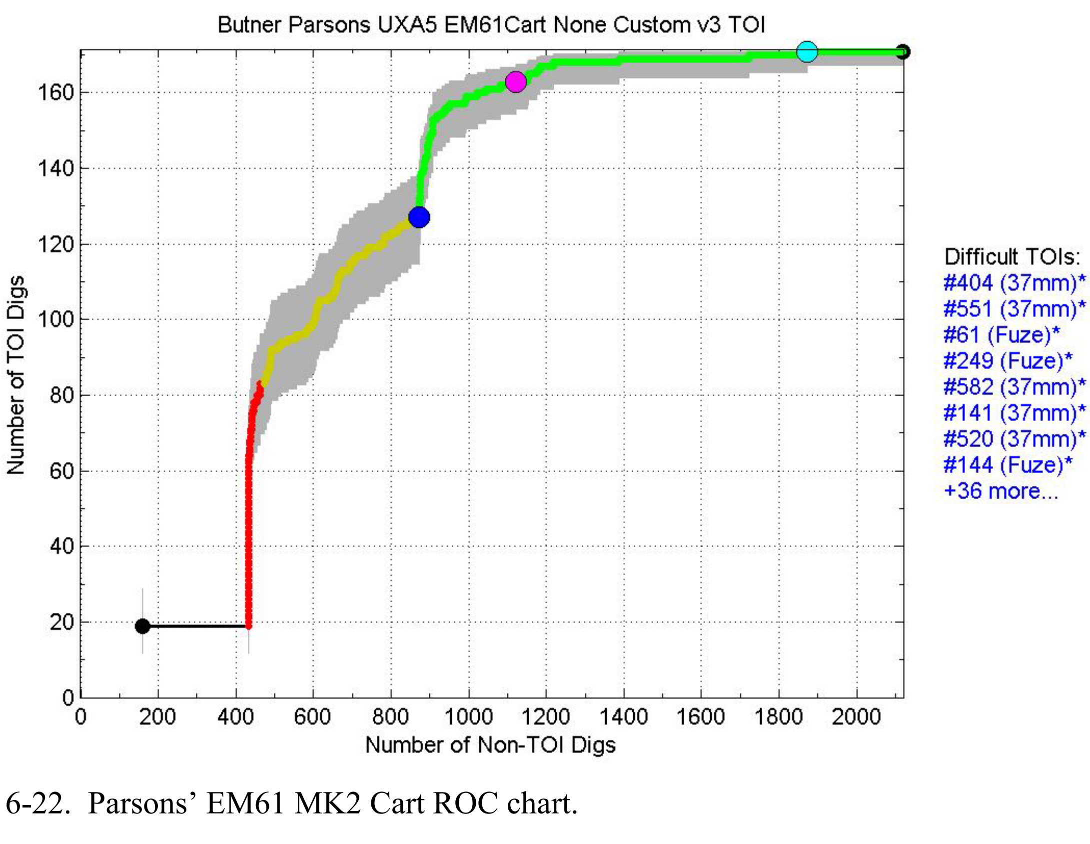


Figure 6-22. Parsons' EM61 MK2 Cart ROC chart.

6.5.3 IDL - TEMTADS

Data from the TEMTADS is shown in Figure 6-23. The crosses identify anomalies selected for Test Set by the ESTCP Program Office.

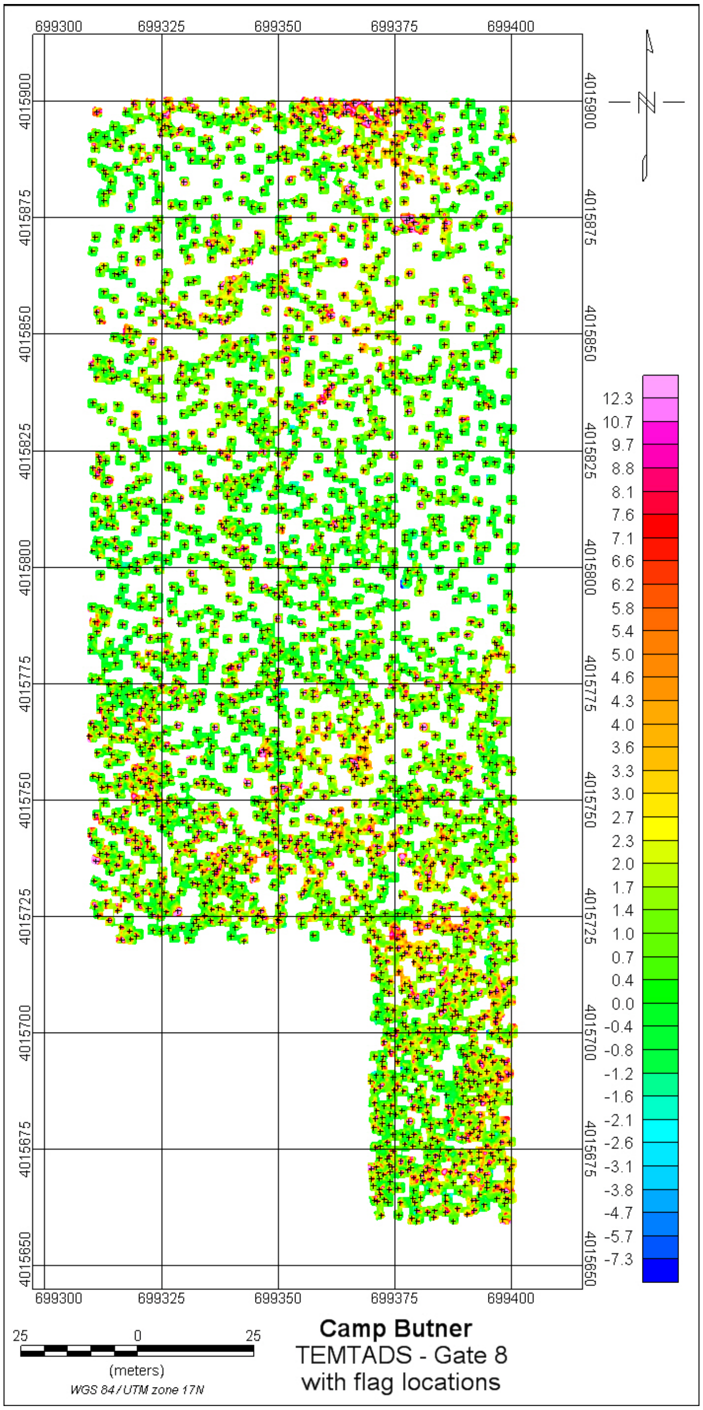


Figure 6-23. TEMTADS data with selected anomalies overlain as crosses.

Performance Scores from IDA

Scoring performances for the NOSLN, 2 and 3 criteria TEMTADS analyses are reported in Table 6-9 to Table 6-12. Their respective ROC charts are shown in Figure 6-24 to Figure 6-27, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 7 false negatives for all the analysis methods. The missed anomalies were nearly identical for all the lists which is not surprising because they used the same target features but with slightly different classification rules.

Table 6-9 Test Set Summary: IDL - TEMTADS – NOSLN 1st pass

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	34	1538	28	7	1607
2	11	493	12	47	563
3	0	1	0	117	118
4	0	2	0	0	2
Training	0	0	0	0	0
TOTAL	**45**	**2034**	**40**	**171**	**2290**

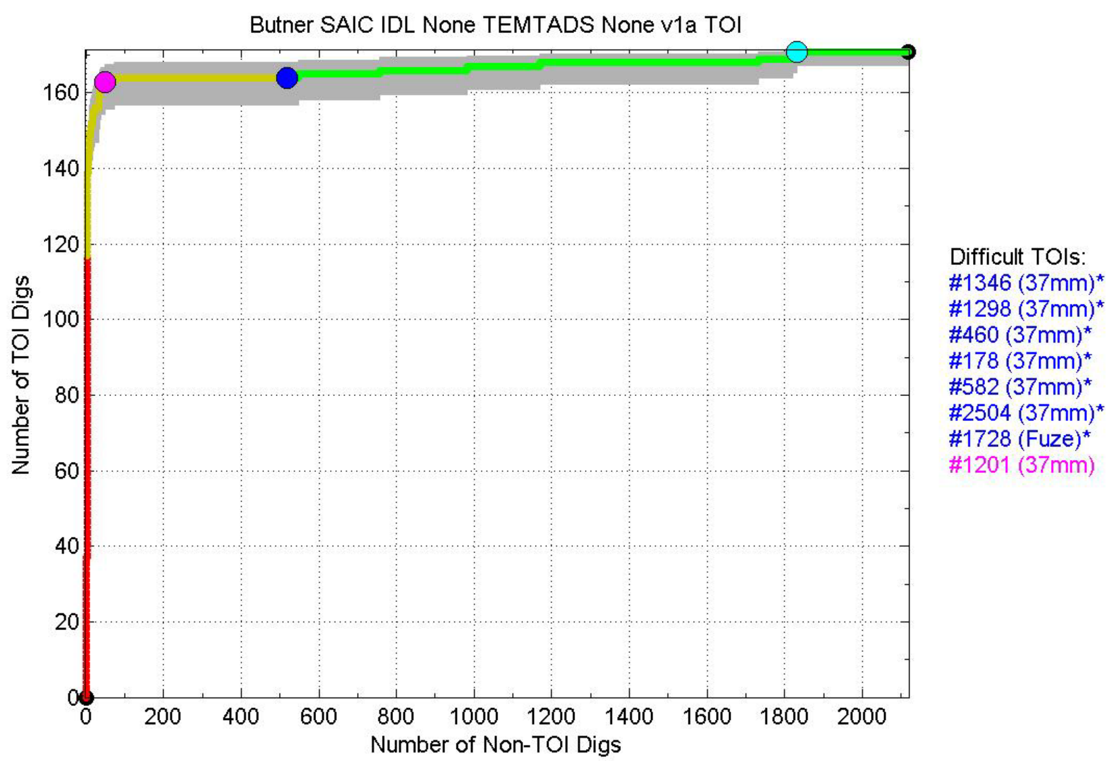


Figure 6-24. IDL - TEMTADS – NOSLN 1st pass ROC chart.

Table 6-10 Test Set Summary: IDL - TEMTADS – NOSLN 2nd pass

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	32	1517	27	7	1583
2	10	464	13	22	509
3	0	3	0	25	28
4	0	2	0	0	2
Training	3	48	0	117	168
TOTAL	**45**	**2034**	**40**	**171**	**2290**

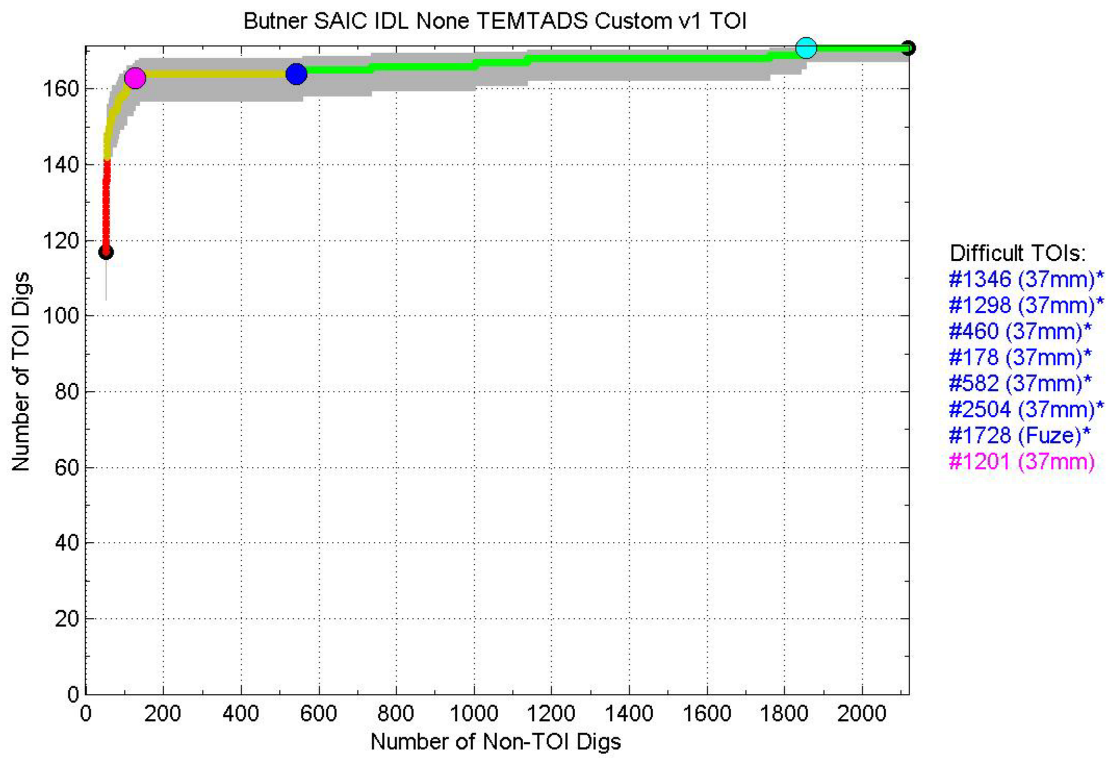


Figure 6-25. IDL - TEMTADS – NOSLN 2nd pass ROC chart.

Table 6-11 Test Set Summary: IDL - TEMTADS – 2-criteria

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	32	1526	28	7	1593
2	7	441	12	0	460
3	1	26	0	135	162
4	0	2	0	0	2
Training	5	39	0	29	73
TOTAL	**45**	**2034**	**40**	**171**	**2290**

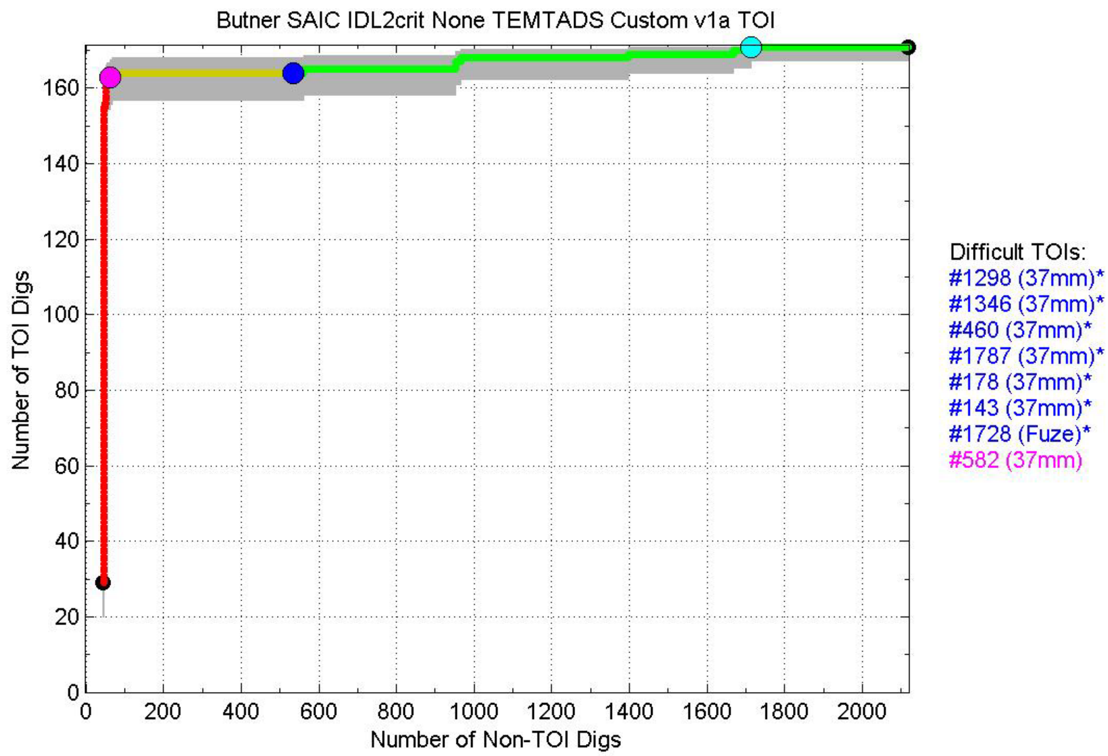


Figure 6-26. IDL - TEMTADS 2-criteria ROC chart.

Table 6-12 Test Set Summary: IDL - TEMTADS – 3-criteria

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	33	1532	28	7	1600
2	7	441	12	0	460
3	0	20	0	135	155
4	0	2	0	0	2
Training	5	39	0	29	73
TOTAL	**45**	**2034**	**40**	**171**	**2290**

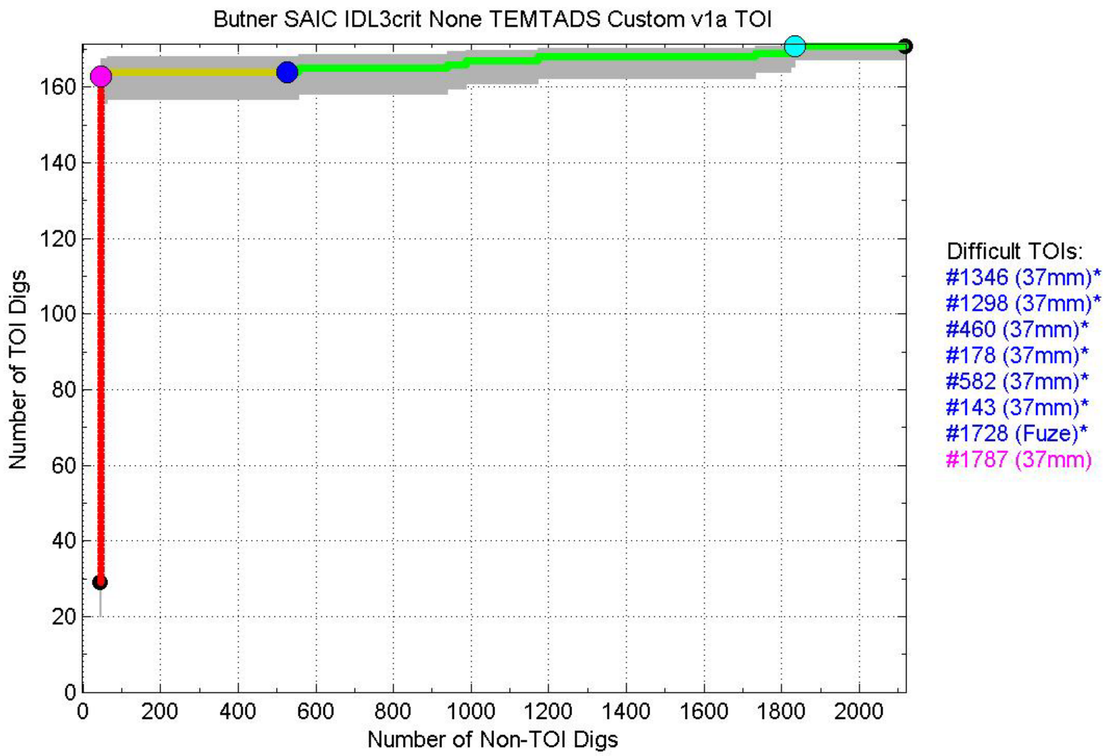


Figure 6-27. IDL - TEMTADS 3-criteria ROC chart.

Characterization Plots

Figure 6-28 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets using the 2-criteria method. The mean error for all TOI with an isolated or slightly overlapping signal was 0.08m with a standard deviation of .04m. If non-TOI are added to the population the mean error increases to 0.11m with a standard deviation of 0.16m.

Figure 6-29 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets using the 2-criteria method. The mean error for all TOI with an isolated or slightly overlapping signal was 0.027m with a standard deviation of .054m. If non-TOI are added to the population the mean error was 0.022m with a standard deviation of 0.047m.

Figure 6-30 show the inverted polarizabilities for all category 1, 2 and 3 targets from the 2-criteria analysis. In general, the high fidelity data obtained by the TEMTADS, as evidenced by the clustering of the TOI in the plots, allows the use of shape to characterize the targets instead of only relying on size which was the case for the EM61 sensors. Table 6-13 tabulates the statistics of the polarizations for the different TOI after removing those anomalies with large fit errors.

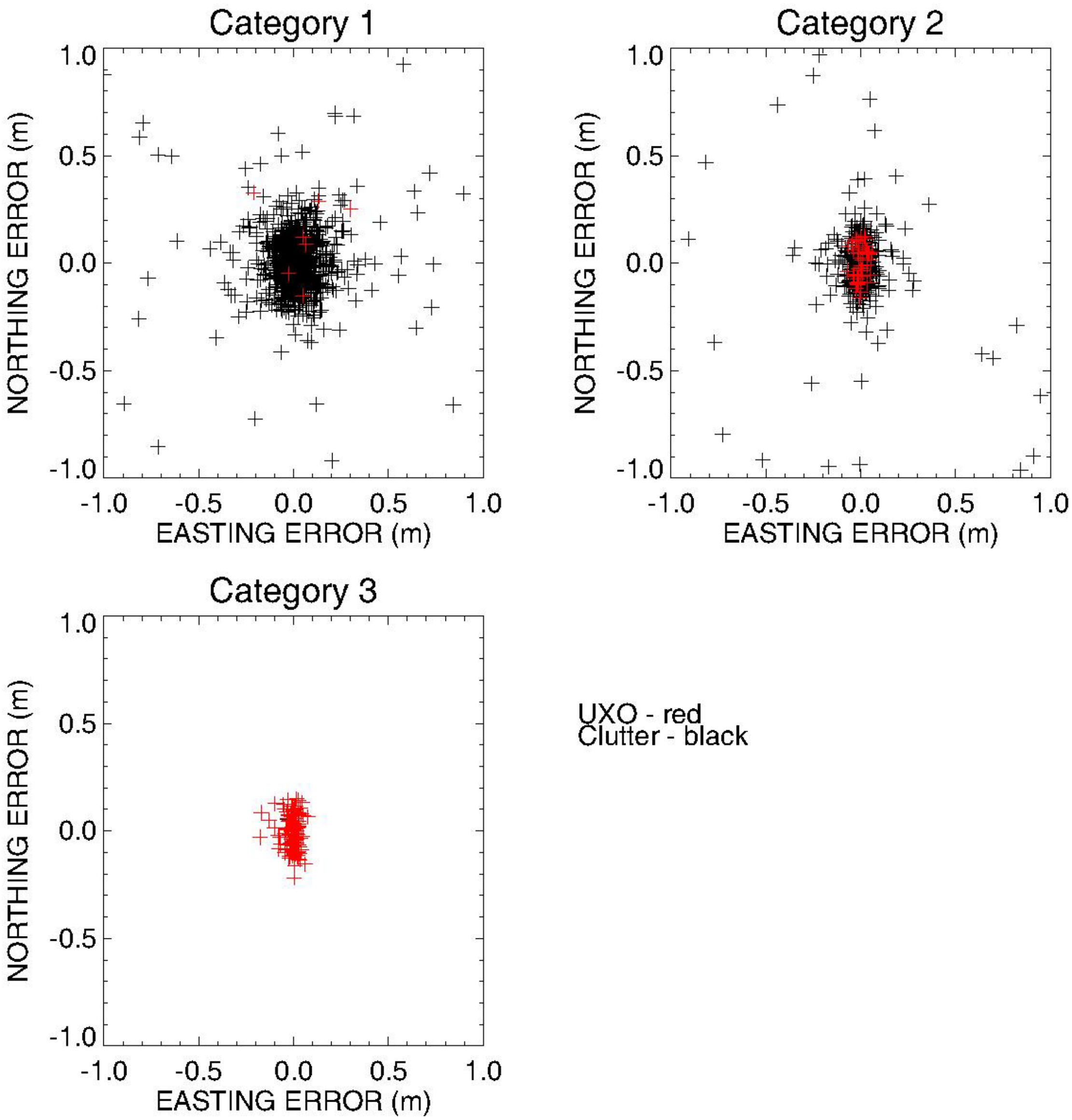


Figure 6-28. Differences between fitted and measured XY locations; TEMTADS 2-criteria analysis

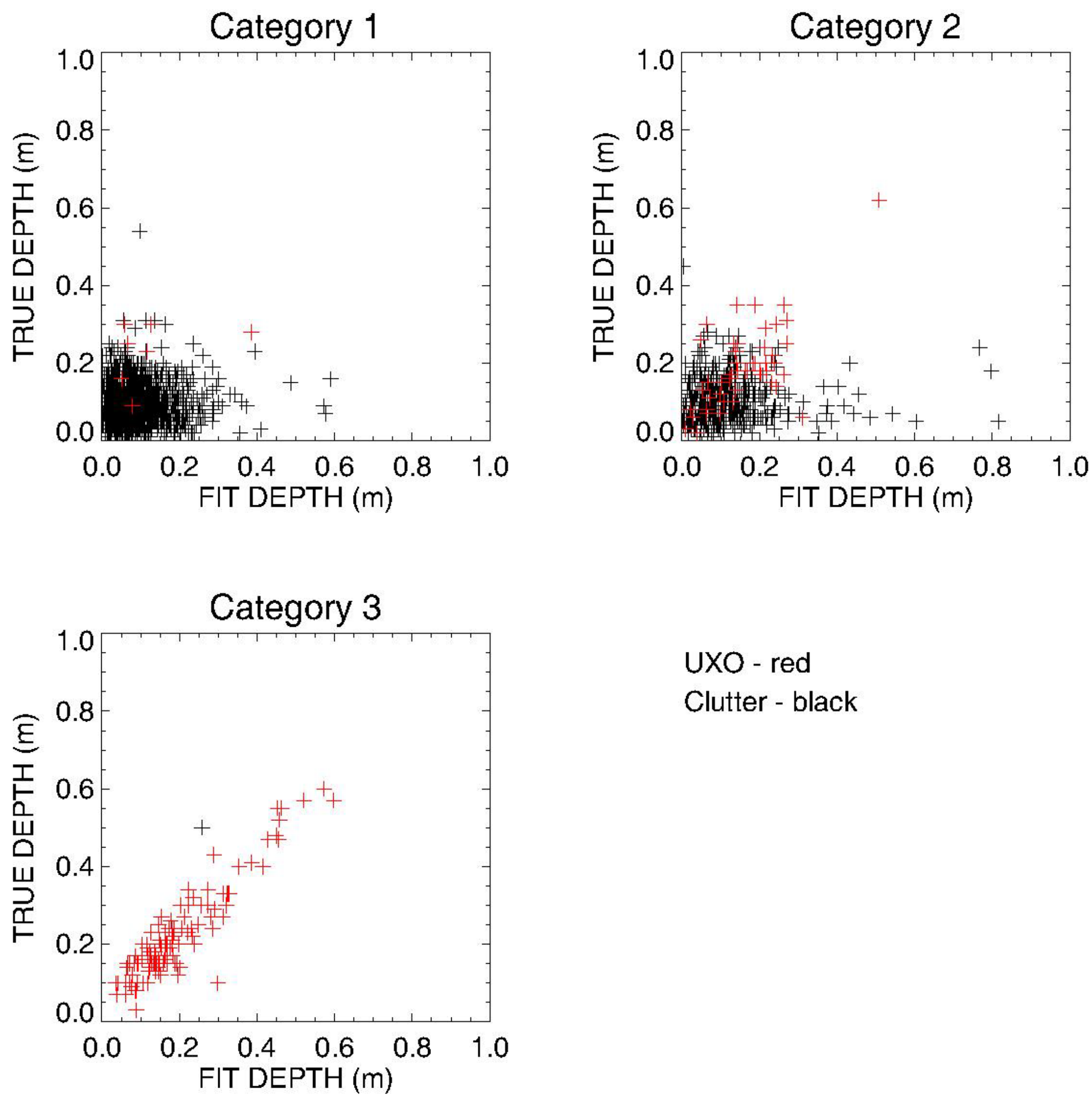


Figure 6-29. Fitted versus measured depth of burial; IDL - TEMTADS 2-criteria analysis

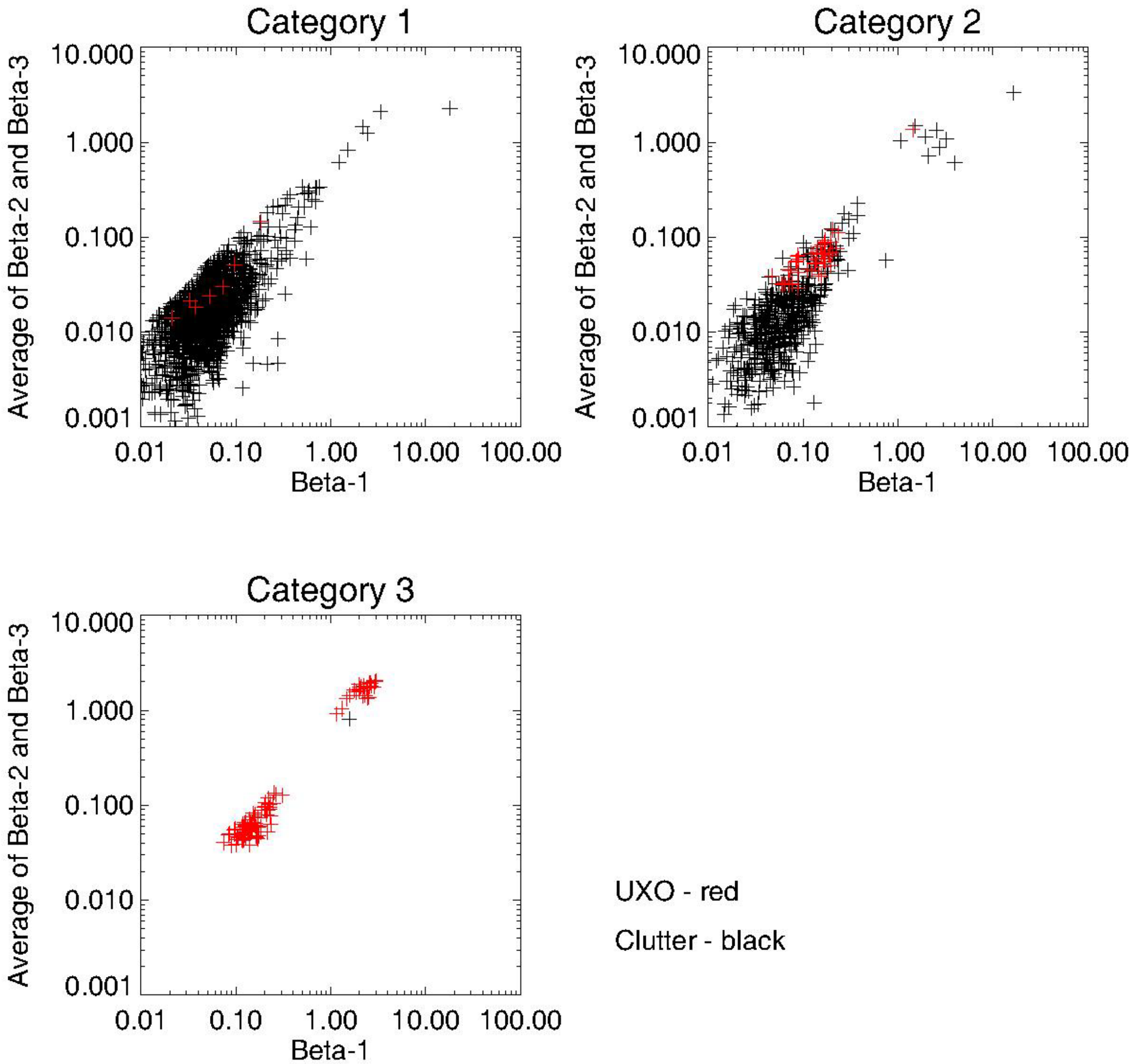


Figure 6-30. Beta 1 versus the average of Beta 2 and Beta 3; IDL-TEMTADS 2-criteria analysis.

Table 6-13 Statistics of betas for the three main TOI, IDL – TEMTADS

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
37mm	100	0.038	0.003	0.135	0.038	0.055	0.013	0.049	0.011
Fuze	22	0.046	0.002	0.216	0.034	0.101	0.017	0.091	0.016
105mm	26	0.107	0.008	2.165	0.504	1.619	0.285	1.418	0.363

Failure Analyses

In Table 6-14, we give the Anomaly IDs of our false negatives for all the analysis methods. Five of the false negatives were on all the lists. All but one of the false negatives was a 37mm.

Table 6-14 TEMTADS' False Negatives

Anomaly ID	NOSLN/3 Criteria/2 Criteria Method	Description
143	2crit/3crit	37mm
178	All	37mm
460	All	37mm
582	NOSLN/3crit	37mm
1298	All	37mm
1346	All	37mm
1728	All	M48 Fuze
1787	2crit	37mm
2504	NOSLN	37mm

Anomalies 1298 and 1346 were missed by all the methods. They were misclassified because the single-dipole solver that was used produced betas that did not match a 37mm and they also failed our axial symmetry test. Figure 6-31 shows the polarization for anomaly 1298. The red and blue lines show polarizations of the field data using two different inversion schemes which we termed eigenvalues and tensor. The eigenvalues solution, shown in red, uses a constant orientation as a function of time gate during the data inversion. This is the inversion scheme we used to produce our target features and decision metrics. The tensor method, shown in blue, allows the orientation to vary as a function of time gate during the inversion. As part of this demonstration, all excavated TOI were measured in air using the TEMTADS. These provide a baseline for the polarizations and are represented on the plot by the black lines. As is evident by the plot the in ground polarizations do not match the ones calculated using in air data. Figure 6-32 shows the results of using the multi-dipole solver on anomaly 1298. The multi-dipole solver found 2 sources and one matched very well to a 37mm in our library. Applying the multi-dipole solver to anomaly 1346 produced similar results. It resulted in four sources of which one was a good match to a 37mm.

During our investigation of the false negatives, it was discovered that both the transmitter and receiver coils in sensor coil #5 of the TEMTADS array were wired with reverse polarity. As sensor #5 is located on the outer edge of the array it typically does not contribute a lot to a given data set if the target is centered on the array. The effects are larger as the targets get closer to the coil. This was the main problem for anomalies 143, 178, 460 and 1787. We applied a correction to the TEMTADS data to fix the polarity of sensor 5 and reran the inversions for the anomalies in question. Figure 6-33 and Figure 6-34 show the polarizations for anomaly 460 as submitted and after fixing the coil. The plots clearly show the improvement in the polarizations after fixing the coil. The other anomalies show similar improvement and their results are tabulated in Table 6-15.

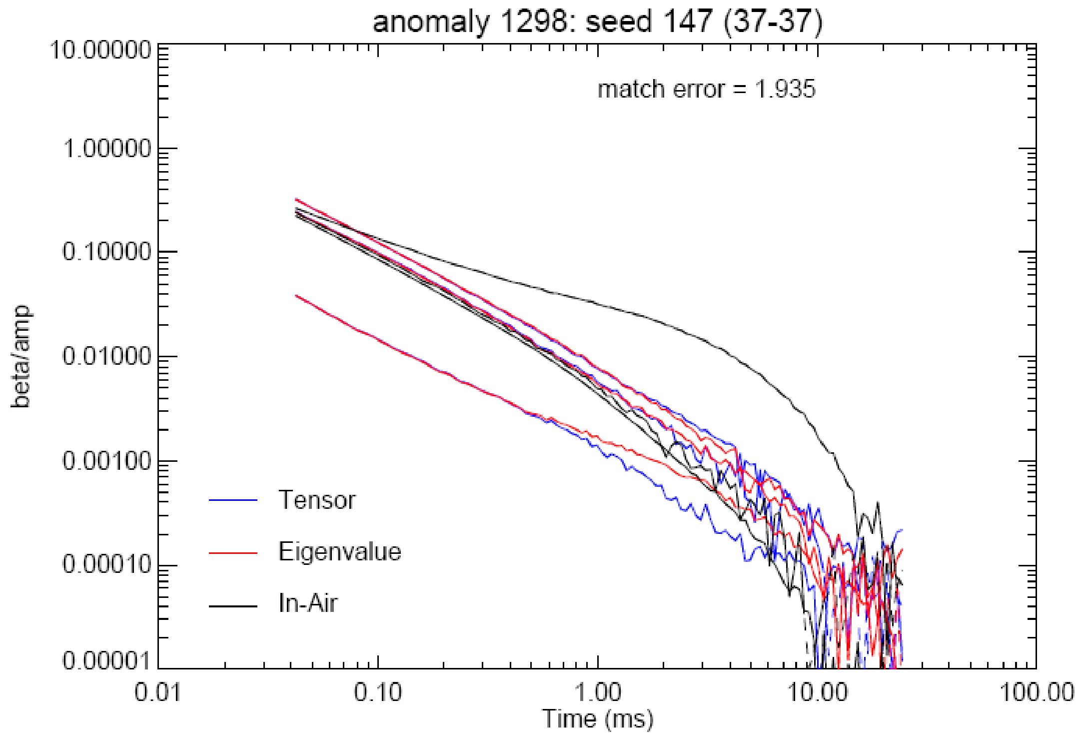


Figure 6-31. TEMTADS polarization plot for anomaly 1298.

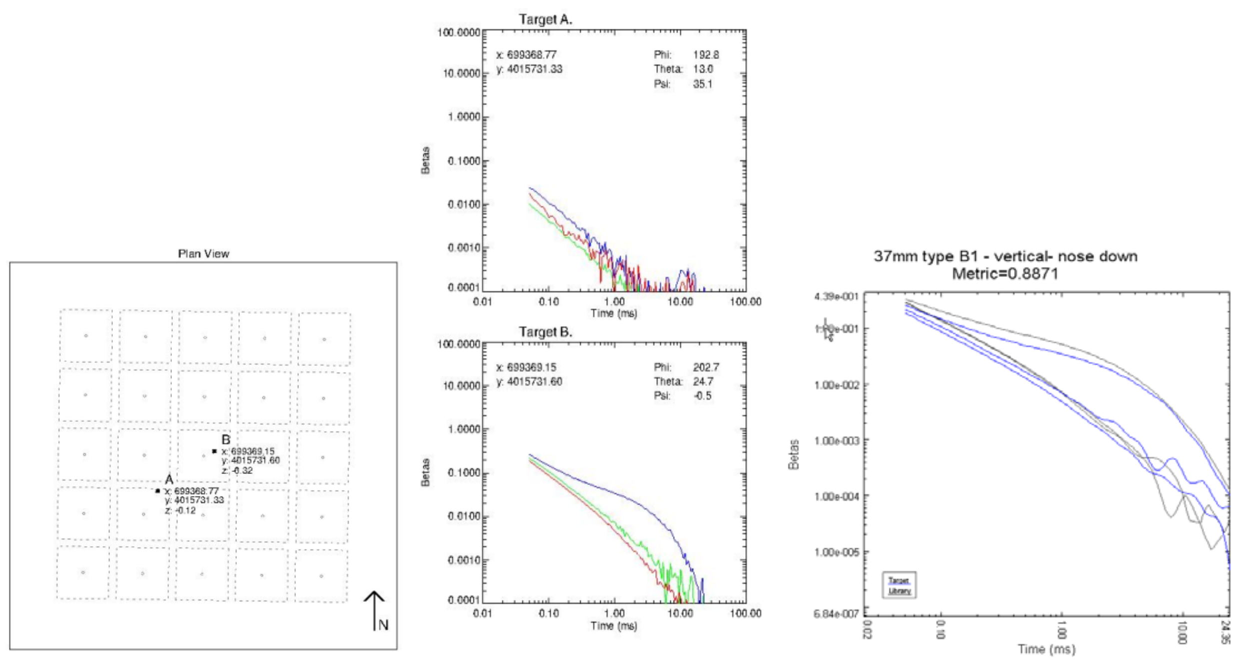


Figure 6-32. Results of running the multi-dipole solver on anomaly 1298.

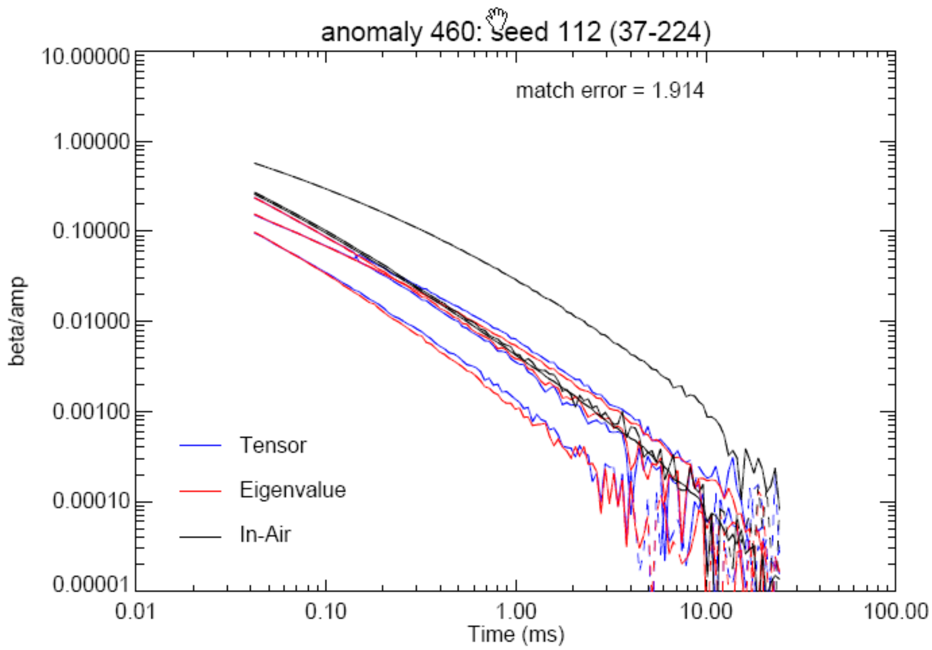


Figure 6-33. TEMTADS polarization plot for anomaly 460

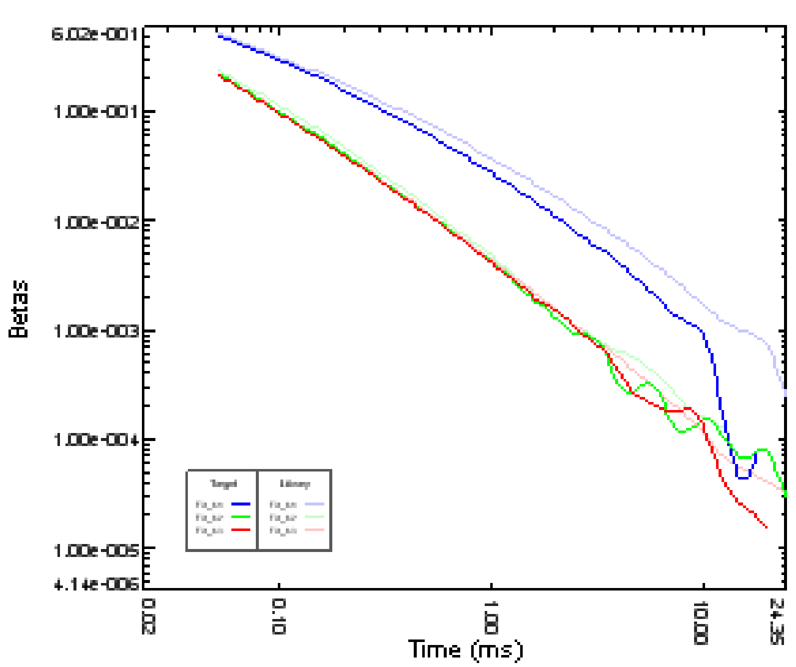


Figure 6-34. TEMTADS polarization plot for anomaly 460 after fixing the polarity of sensor #5.

Table 6-15. TEMTADS 2-criteria and 3-criteria metrics before and after fixing the polarity of coil #5.

Anomaly ID	2 crit - submitted	2 crit after fix	3 crit - submitted	3 crit after fix
143	0.7690	0.8559	0.6709	0.7309
178	0.7690	0.9617	0.4657	0.7934
460	0.4544	0.8986	0.3344	0.8610
1787	0.7387	0.9331	0.7010	0.8680

The remaining false negatives (582, 1728 and 2504) were caused by a combination of the aforementioned problems. By fixing the polarity problem with sensor 5 followed by inversion using the multi-dipole solver we are able to retrieve polarizations that match with TOI in the library. This is illustrated for anomaly 582 in Figure 6-35 and Figure 6-36 where the first figure shows the polarization as submitted and second figures shows the results of the multi-dipole solver on the corrected data. This produced a good match to a 37mm.

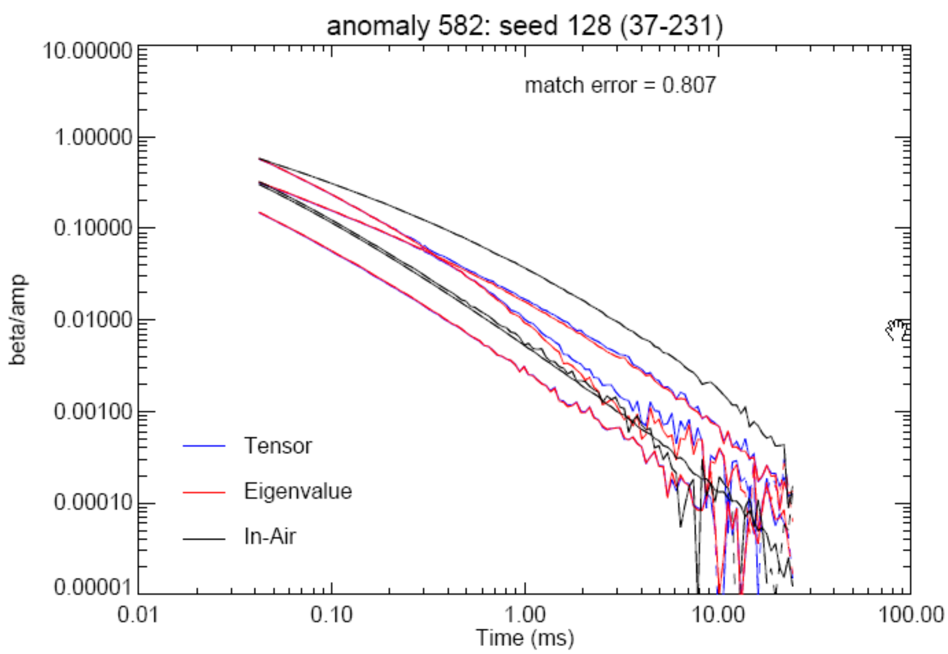


Figure 6-35. TEMTADS polarization plot for anomaly 582.

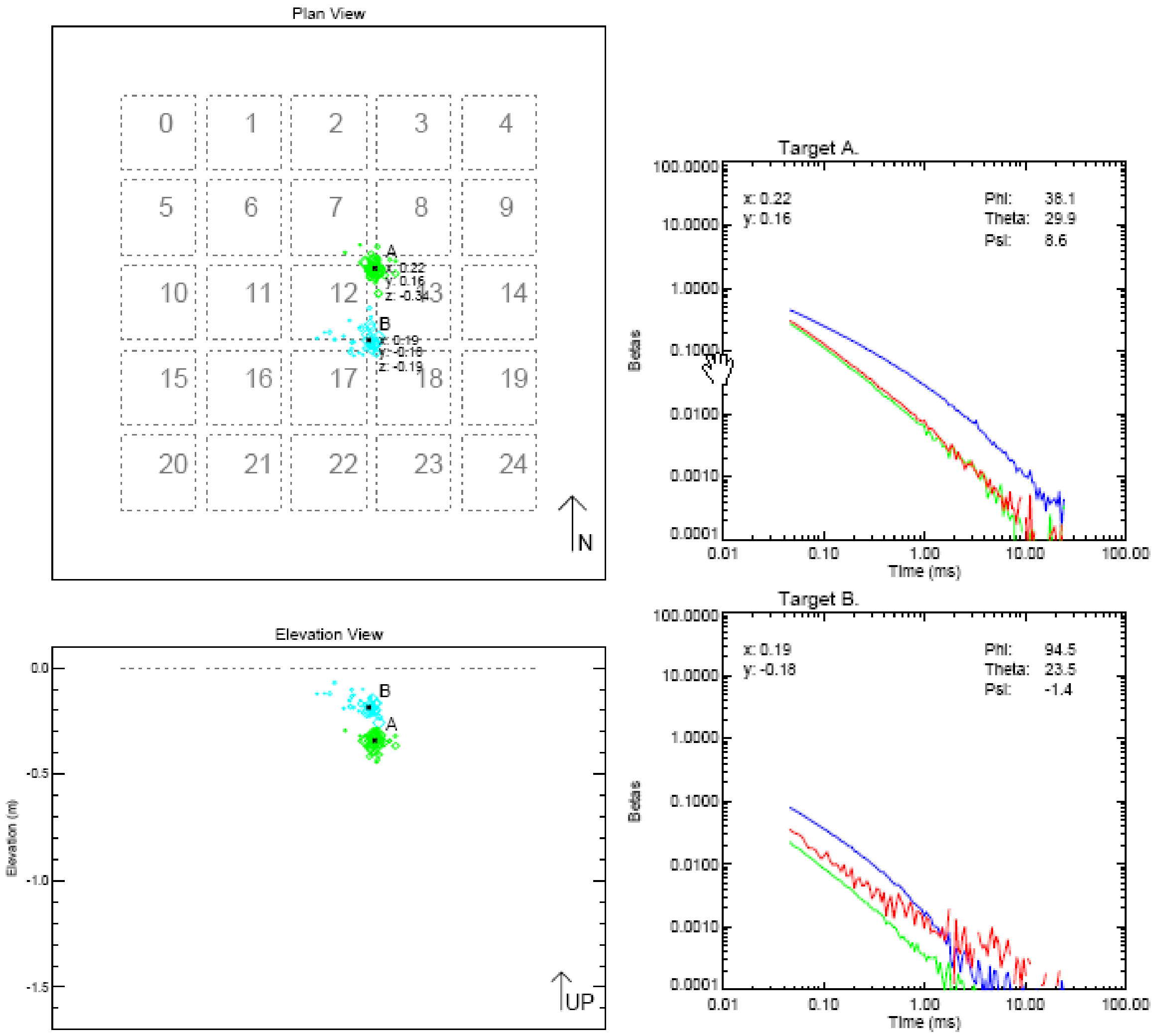


Figure 6-36. Results of fixing the polarity of sensor coil #5 and running the multi-dipole solver on anomaly 582.

6.5.4 UX-ANALYZE - TEMTADS

Performance Scores from IDA

Scoring performances for the SAIC's UX-Analyze TEMTADS analysis using no training and the 2-criteria metric are reported in Table 6-16. The respective ROC chart is shown in Figure 6-37, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 3 false negatives and although not a false negative, anomaly 460 was close to the dig/no dig threshold with the next TOI being over 1000 digs away. Anomaly IDs of the false negatives were 429, 1201 and 2504.

Table 6-16 Test Set Summary: UX-Analyze - TEMTADS

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	24	835	18	3	880
2	19	1199	22	29	1269
3	0	2	0	139	141
4	0	0	0	0	0
Training	0	0	0	0	0
TOTAL	**43**	**2036**	**40**	**171**	**2290**

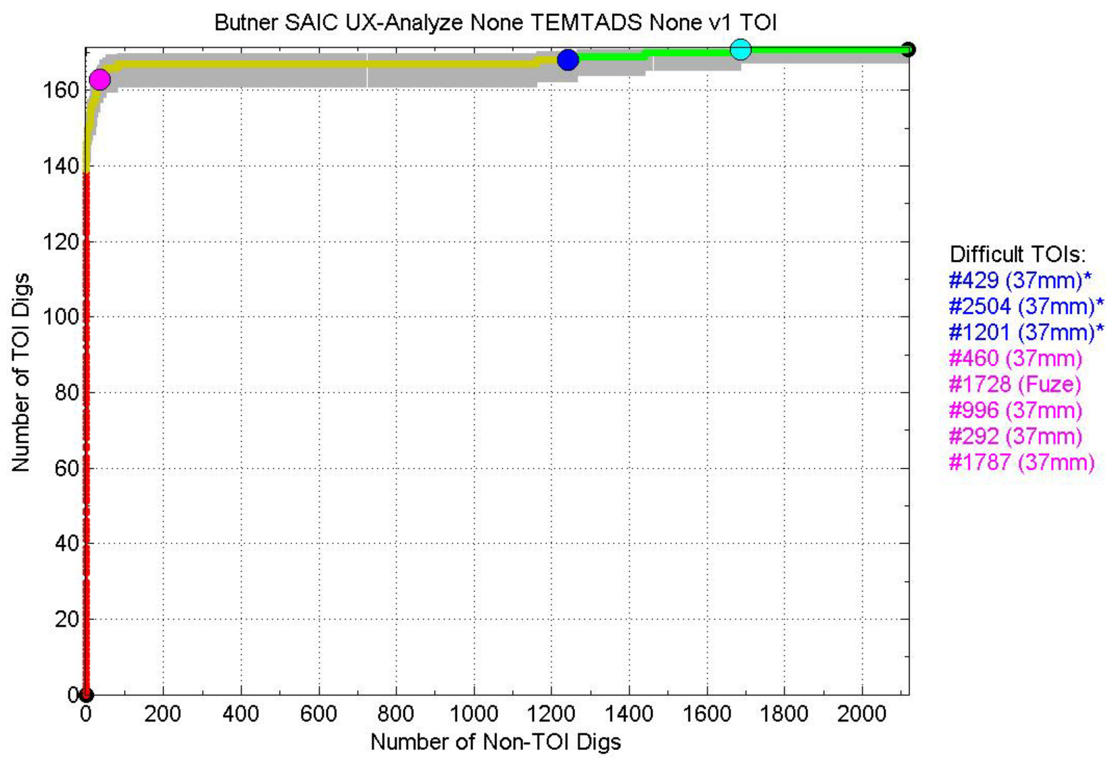


Figure 6-37. UX-Analyze - TEMTADS ROC chart.

Characterization Plots

Figure 6-38 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets. The mean error for all TOI with an isolated or slightly overlapping signal was 0.10m with a standard deviation of .054m. If non-TOI are added to the population the mean error increases to 0.23m with a standard deviation of 0.28m.

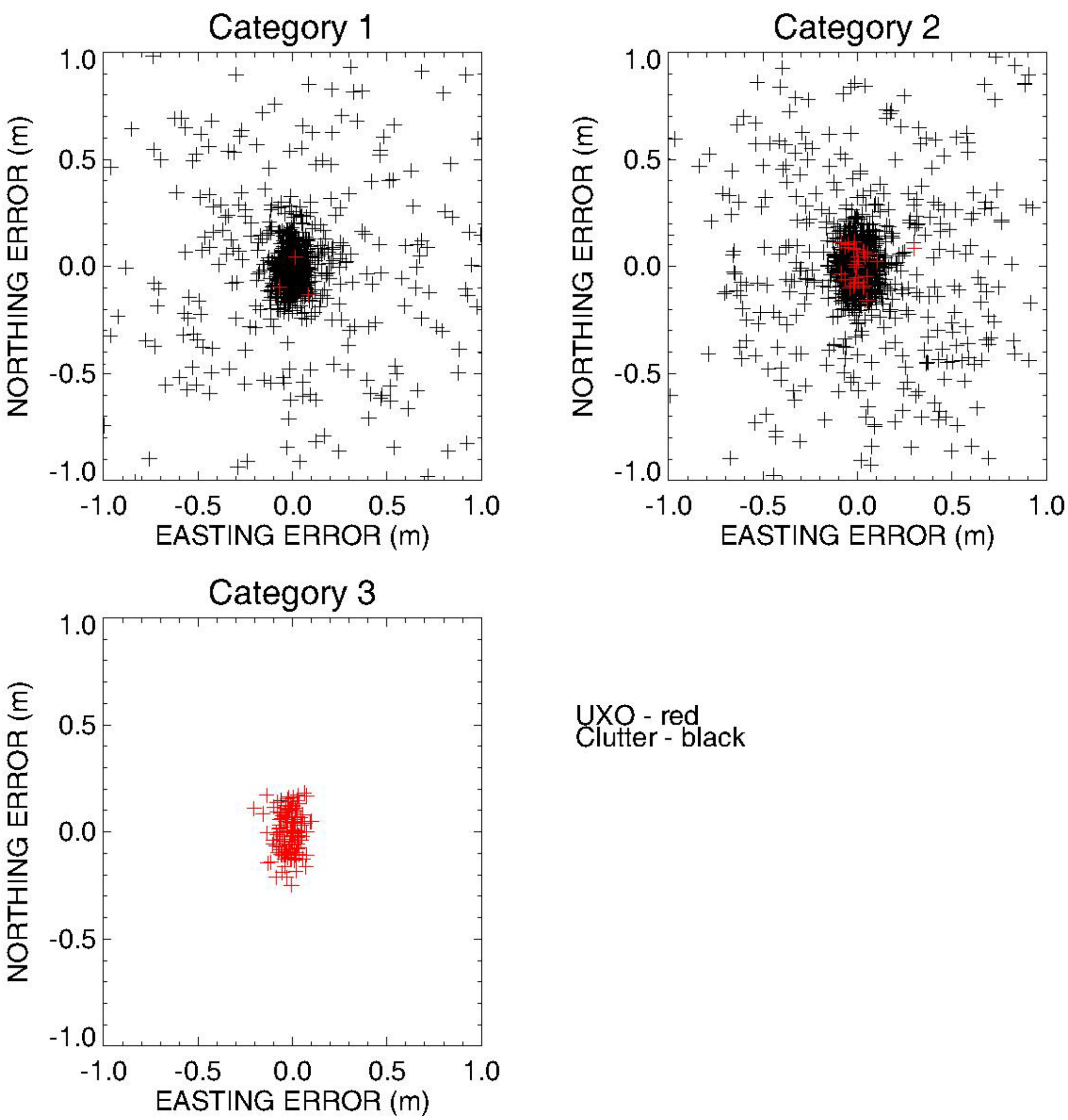


Figure 6-38. Differences between fitted and measured XY locations; UX-Analyze TEMTADS 2 criteria analysis

Figure 6-39 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets. The mean error for all TOI with an isolated or slightly overlapping signal was 0.029m with a standard deviation of .053m. If non-TOI are added to the population the mean error was 0.023m with a standard deviation of 0.055m.

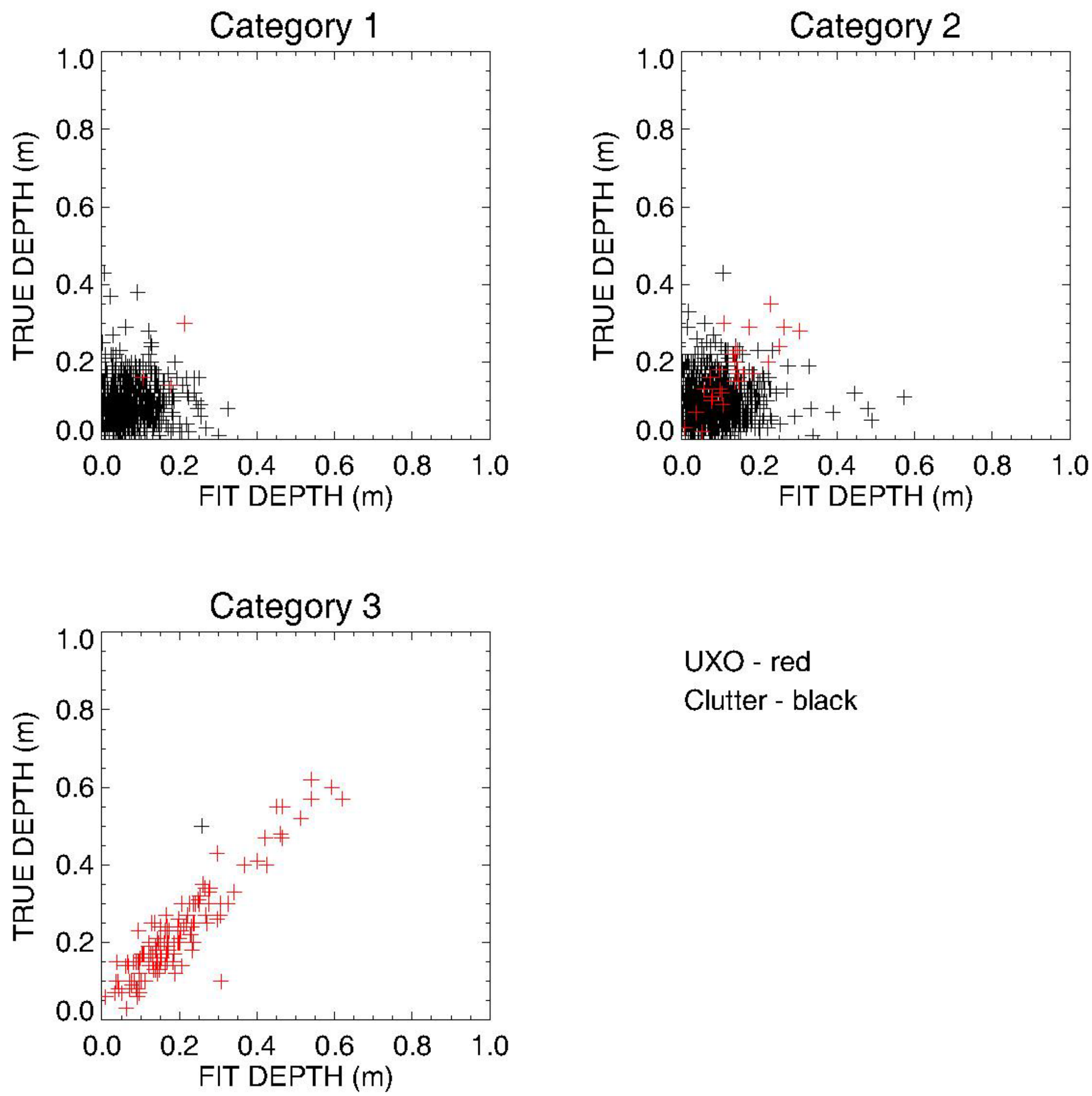


Figure 6-39. Fitted versus measured depth of burial; UX-Analyze - TEMTADS 2 criteria analysis

Figure 6-40 show the inverted polarizabilities for all category 1, 2 and 3 targets. As with the IDL-TEMTADS characterization plots, there is good clustering of the TOI which allows the use of shape to characterize the targets. Table 6-17 tabulates the statistics of the polarizations for the different TOI after removing those anomalies with large fit errors.

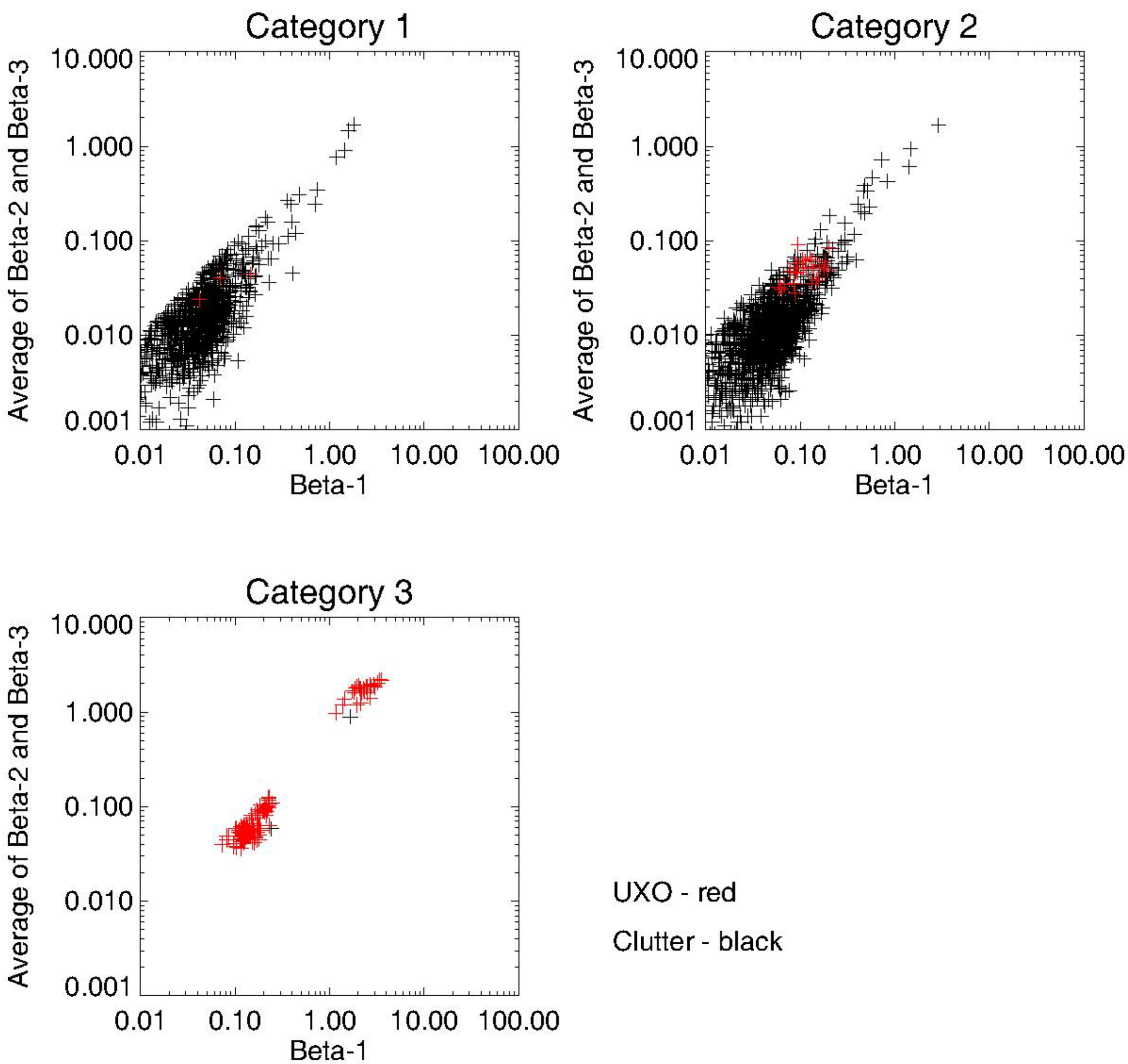


Figure 6-40. Beta 1 versus the average of Beta 2 and Beta 3; UX-Analyze TEMTADS 2 criteria analysis.

Table 6-17 Statistics of Betas for the three main TOI, UX-Analyze – TEMTADS

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
37mm	106	0.038	0.003	0.128	0.032	0.052	0.010	0.045	0.009
Fuze	21	0.045	0.002	0.208	0.023	0.095	0.011	0.086	0.008
105mm	24	0.109	0.009	2.335	0.621	1.689	0.319	1.535	0.406

Failure Analyses

The main cause of the all the false negatives and the late Category 2 anomaly was the polarity problem of sensor coil #5 discussed in the previous section. The UX-Analyze analysis used the multi-dipole solver and the single-dipole solver and retained the results which were closest to a TOI. Anomaly 2504 was the one anomaly that needed the multi-dipole solver in conjunction with fixing the polarity problem to produce polarization that matched well with the TOI in our library. The other three anomalies needed only to fix the polarity problem because both the single and multi-dipole solvers produced good matches to a TOI. Anomaly 460 was discussed in the previous section and produced similar results in UX-Analyze. Figure 6-41 shows the best two library matches for anomaly 1201 from our initial analysis using the single-dipole solver. The best metric was 0.7118 which was below our threshold of 0.81. The second best match had a metric of 0.5815 but visual inspection of the plots shows that the decay of the polarizations match well to the 37mm but are shifted in amplitude. This amplitude shift caused the lower metric. Figure 6-42 shows the two best matches to our library after fixing the polarity problem. The results shown also used the single-dipole solver. The metric is now 0.8229 which would have exceeded our cutoff and visual inspection of the plots clearly show polarizations with a better match to a 37mm.

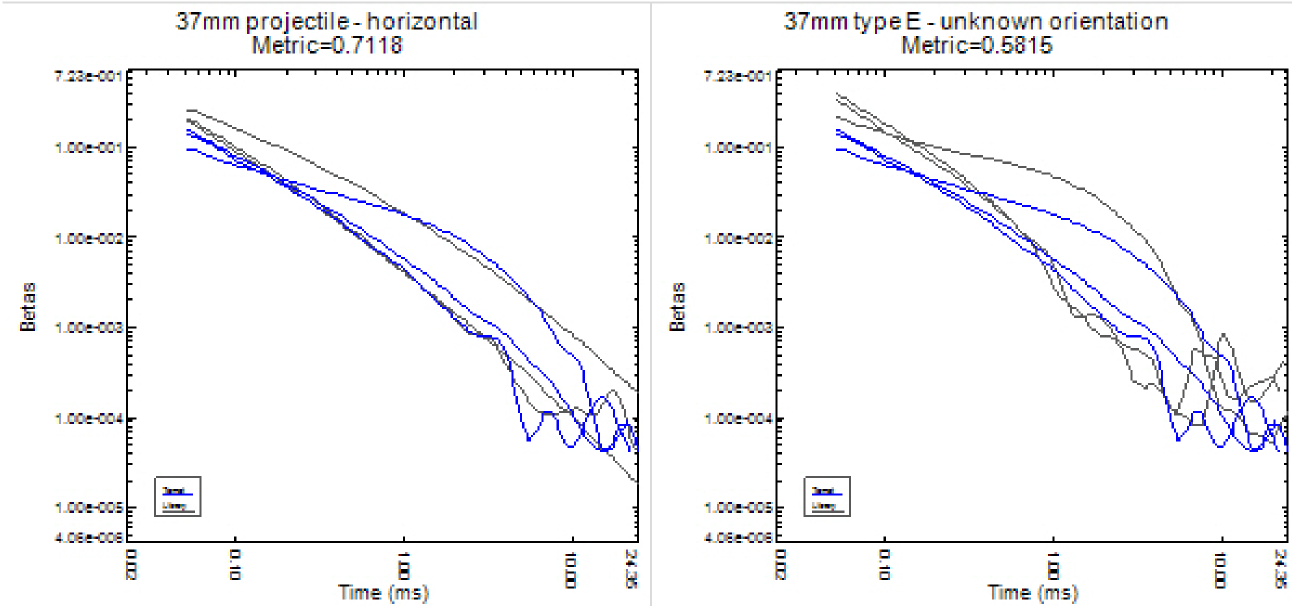


Figure 6-41. Two best library matches for anomaly 1201 using TEMTADS single-dipole solver.

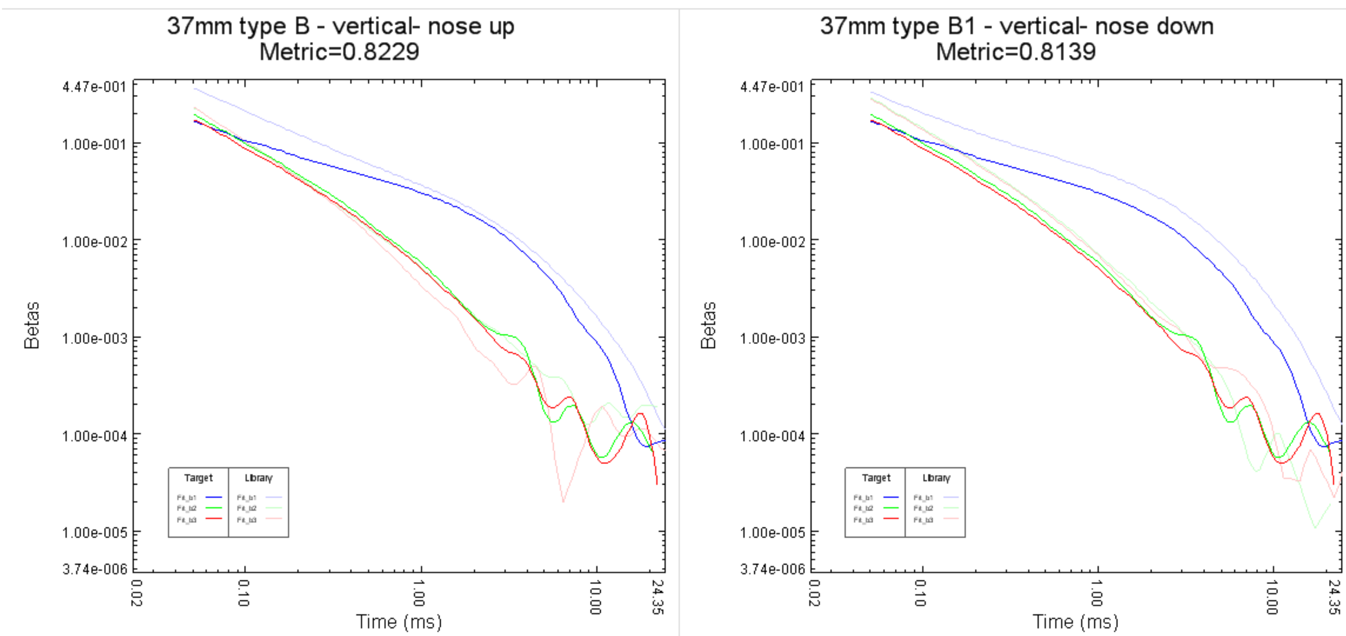


Figure 6-42. Two best library matches for anomaly 1201 after fixing the polarity problem for sensor #5.

6.5.5 UXANALYZE - EM61 MK2 CART with TEMTADS

The EM61-MK2 data was used as a first pass to try to discriminate between TOI and clutter. The EM61-MK2 pre-screener classified 319 out of 2290 anomalies with the remaining 1971 requiring TEMTADS cued data. The specific breakdown was 68 Category 3 anomalies and 251 Category 1 anomalies. These anomalies were place at the beginning and end of our dig list and bookended the TEMTADS results.

Performance Scores from IDA

Scoring performances for the UX-Analyze EM61-MK2/TEMTADS analysis are reported in Table 6-18. A ROC chart is shown in Figure 6-43, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 9 false negatives. Anomalies 61, 404, 543, 429, 582, 178, 2504, 1201 and 1981 were all classified as high confidence clutter. Two of the false negatives were the result of the EM61 pre-screener.

Table 6-18 Test Set Summary: Non SCORR EM61 MK2 Cart

Category	Cultural	Munition Debris	No Contact	Soil	UXO	TOTAL
1	28	1083	31	0	9	1151
2	16	942	9	0	31	998
3	1	9	0	0	131	141
4	0	0	0	0	0	0
Training	0	0	0	0	0	0
TOTAL	**45**	**2034**	**40**	**0**	**171**	**2290**

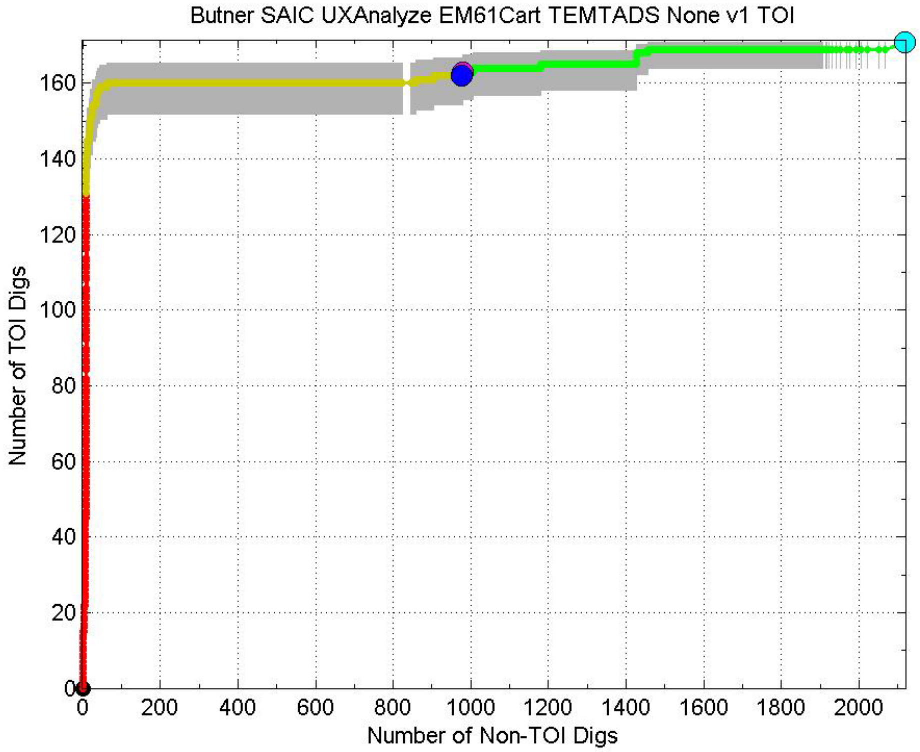


Figure 6-43. UX-Analyze EM61 MK2 Cart with TEMTADS ROC chart.

Failure Analyses

Anomalies 61 and 404 were classified as high confidence clutter using the EM61 data. The cause of the failure for anomaly 404 is the same as discussed in the section describing the EM61 only data. The TAU14 parameter for anomaly 61 was 474 which was slightly smaller than the 487 cutoff but had a standard deviation of 62 which is quite high. Figure 6-44 shows the measured and modeled data for anomaly 61. The anomaly shape is irregular and in combination with the high standard deviation for the TAU14 decay parameter the anomaly should have been tagged as overlapping and a candidate for cued data.

The seven false negatives that were due to analysis of the TEMTADS data were caused by the polarity problem with sensor #5. They consisted of the same three false negatives as the UX-Analyze TEMTADS only analysis and anomalies 178, 543, 582 and 1981. Many of the anomalies in question have been discussed in the previous sections and the cause of their failures in this analysis were the same.

The reason for the additional failures with this method was that only the multi-dipole solver was used whereas the TEMTADS only analysis also used the single-dipole solver. The multi-dipole solver appears to be more sensitive to the polarity problem with coil #5 and resulted in a slightly

lower match metric. The library metric for these failures ranged from 0.668 to 0.803 which are close to our threshold of 0.81. Lowering our threshold would have also averted the false negatives without adding many false positives but the most efficient solution would be to fix the polarity problem. Figure 6-45 shows the best two library matches for anomaly 543 from our initial analysis with the best metric being 0.668. Similar to anomaly 1201, visual inspection of the plots shows that the decay of the polarizations match well to a 37mm but are shifted in amplitude, especially the primary polarization. Figure 6-46 shows the two best matches to our library after fixing the polarity problem. The amplitude shift problem has been resolved resulting in a library metric of 0.9396 which is a very good match to a 37mm. Applying the same process to anomaly 1981 resulted in a similar improvement to the library metric.

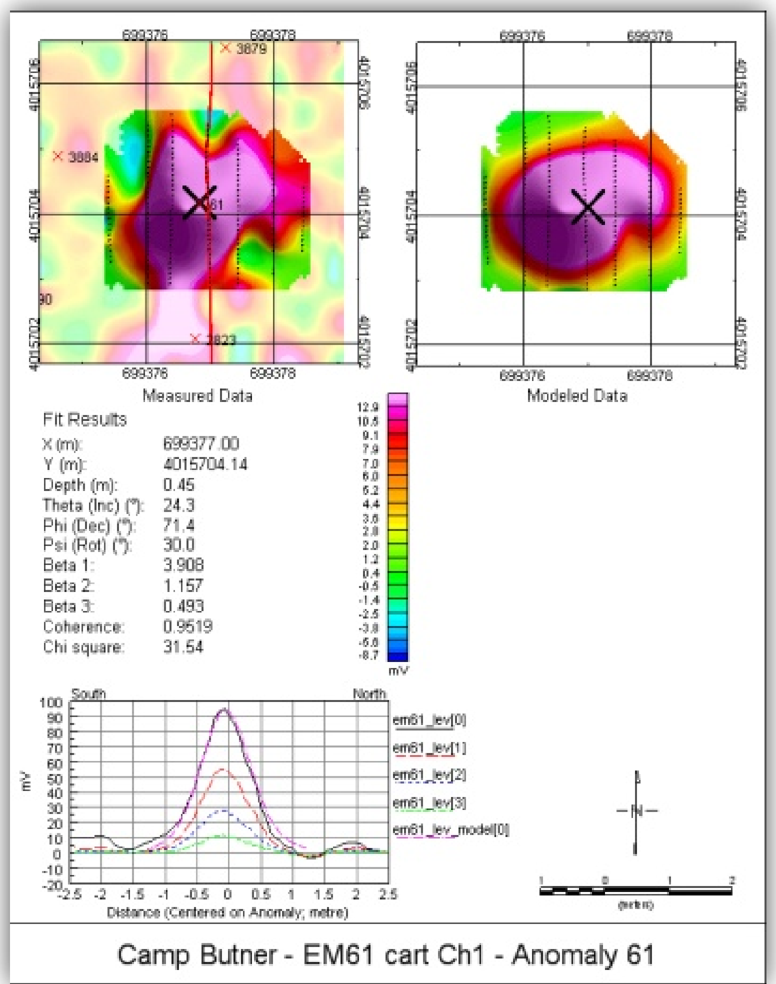


Figure 6-44. Anomaly plot showing measured data, inverted features and forward model for anomaly 61 which was incorrectly ranked as high confidence clutter.

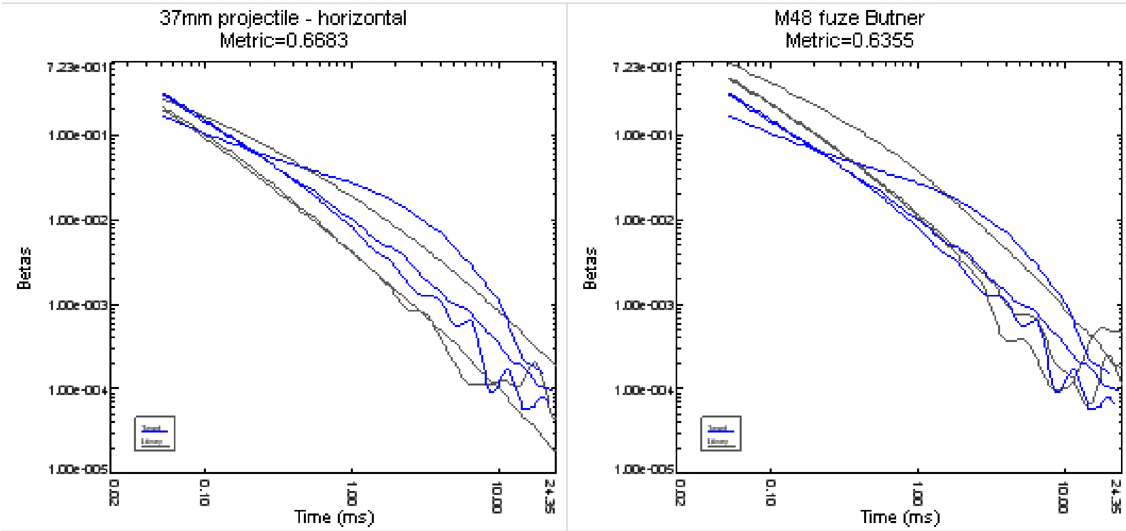


Figure 6-45. Two best library matches for anomaly 543 using TEMTADS single-dipole solver.

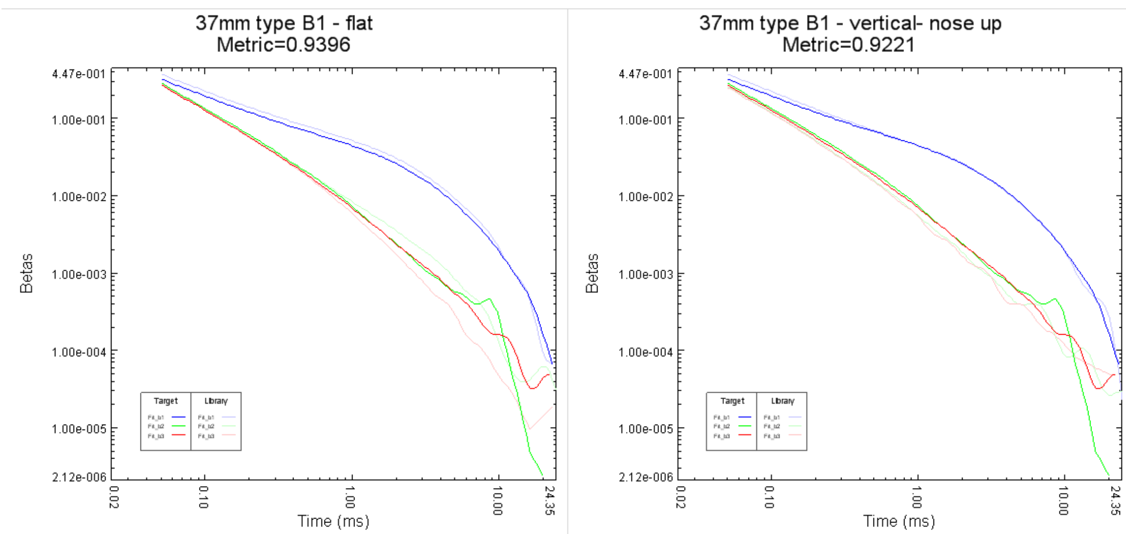


Figure 6-46. Two best library matches for anomaly 543 after fixing the polarity problem for sensor coil #5.

6.5.6 UXANALYZE - EM61 MK2 CART with METAL MAPPER

The EM61-MK2 data was used as a first pass to try to discriminate between TOI and clutter. The EM61-MK2 pre-screener was identical to the one describe in a Section 6.5.5. It classified 319 out of 2290 anomalies with the remaining 1971 requiring MM cued data.

Performance Scores from IDA

Scoring performances for the UX-Analyze EM61-MK2/MM analysis are reported in Table 6-19. A ROC chart is shown in Figure 6-43, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 8 false negatives. Similar to the EM61-MK2/TEMTADS analysis, anomalies 61 and 404 were classified as high confidence clutter based on the EM61-MK2 data. Anomalies 884, 1154, 1298, 1346, 2340 and 2504 were false negatives using the MM data.

Table 6-19 Anomaly Summary: UX-Analyze - EM61-MK2 Cart/MM

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	25	729	28	8	790
2	15	1185	12	7	1219
3	5	104	0	156	265
4	0	16	0	0	16
Training	0	0	0	0	0
TOTAL	**45**	**2034**	**40**	**171**	**2290**

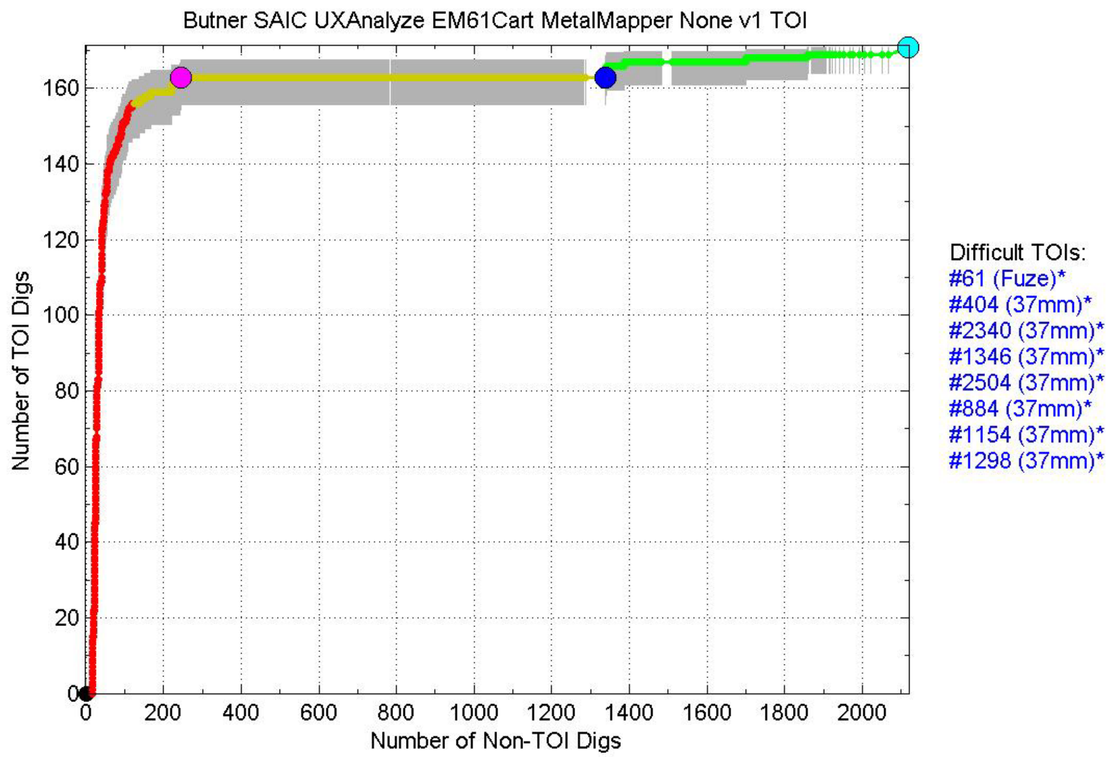


Figure 6-47. UX-Analyze EM61-MK2 Cart with Metal Mapper ROC chart.

Failure Analyses

Anomalies 61 and 404 were classified as high confidence clutter using the EM61 data. The cause of these failures was discussed in Sections 6.5.1 and 6.5.5.

Almost all of the false negatives resulted from data collected using the Geometrics MM. Anomaly 1346 was the only one that had data from the Sky system but this anomaly also had data from the Geometrics system.

Four of the false negatives (anomalies 884, 1154, 1298 and 2340) were caused by bad data from receiver coil 3Y which is the center coil on the MM. The UX-Analyze analysis team was unaware of intermittent problems with this coil on the Geometrics MM. Figure 6-48 shows a QC plot of the MM data for anomaly 2340 with the problematic Rx3Y data circled. The figure is organized in three row based sections each corresponding to a transmitter. The first three rows are associated with data when the X transmitter is fired. The middle three rows present the Y transmitter data and the bottom 3 rows the Z transmitter data. Within each section the three rows present the X, Y and Z axis receiver data. The columns indicated the receiver coil number going from 0 to 6 as we move left to right. All the data plotted as red lines are negative while the black lines show positive responses. Figure 6-49 shows the inverted results including Rx3Y (left) and after removing it (right). The new polarizations produced without the Rx3Y data match well to a

37mm. The polarizations for the three other anomalies also matched well to 37mms after we removed the Rx3Y data.

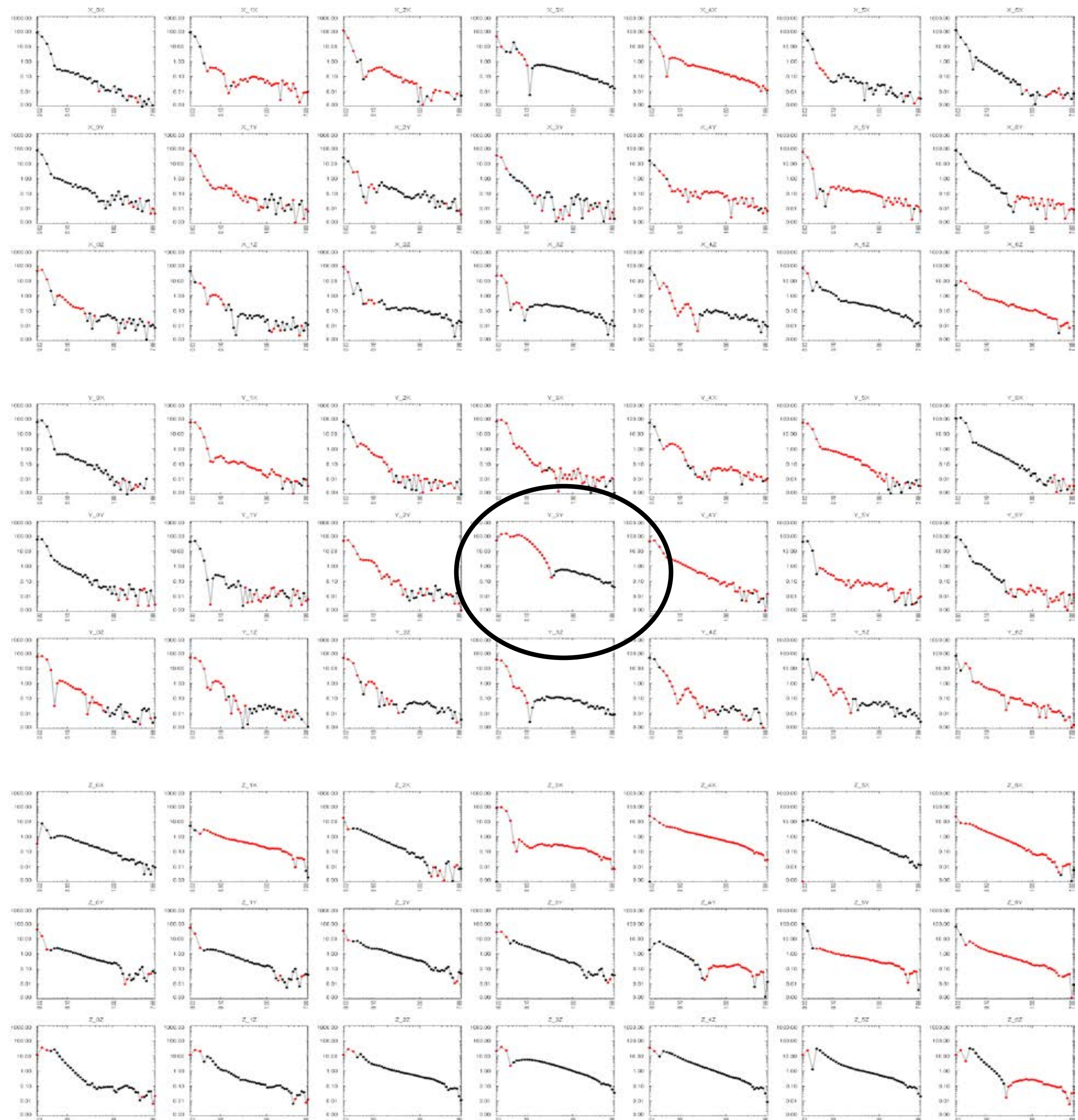


Figure 6-48. Measured decay QC plot of the MM data for anomaly 2340 with the problematic Rx3Y data circled.

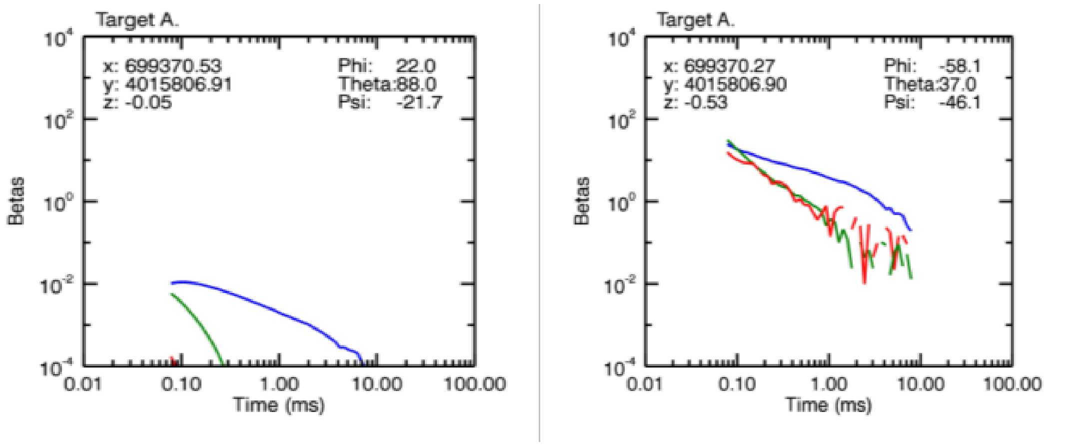


Figure 6-49. The results of the inversion of MM data for anomaly 2340 with Rx3Y (left) and without (right).

Anomaly 2504 inverted to yield very unusual betas, as shown in Figure 6-50. The left plot shows the polarizations using the single-dipole solver while the multi-dipole results are on the right. Figure 6-51 shows the polarizations of the single and multi dipole solvers after removing the Rx3Y data. These polarizations visually appear more normal but both have fit model errors exceeding 85% which are much greater than the desired 10% or less. The peak amplitude for this anomaly was 2.4μT/A-s which was just above the low amplitude threshold of 2.04μT/A-s. The combination of low amplitude and high fit error should have classified this anomaly as a Category 2 or “Cannot decide”.

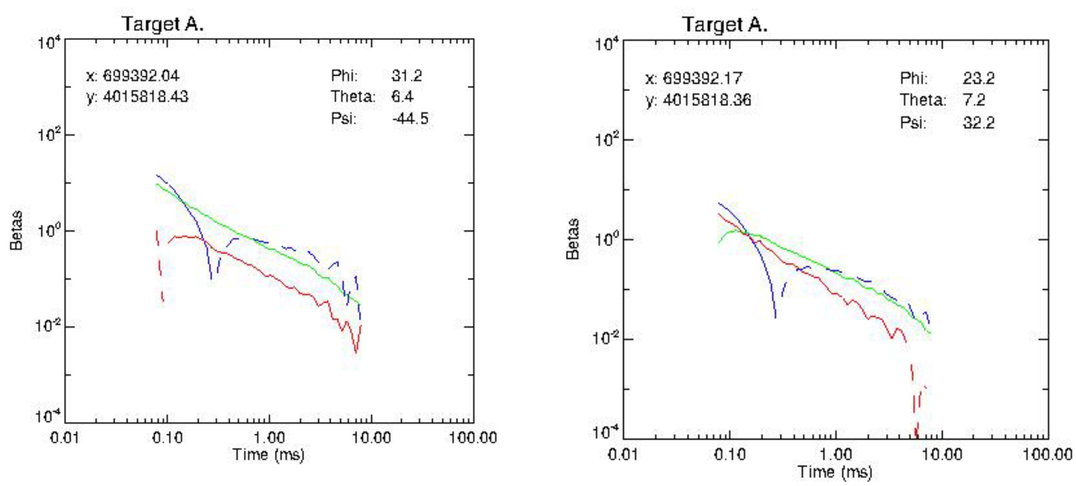


Figure 6-50. Polarization plots of anomaly 2504 using the single-dipole solver (left) and multi-dipole solver (right).

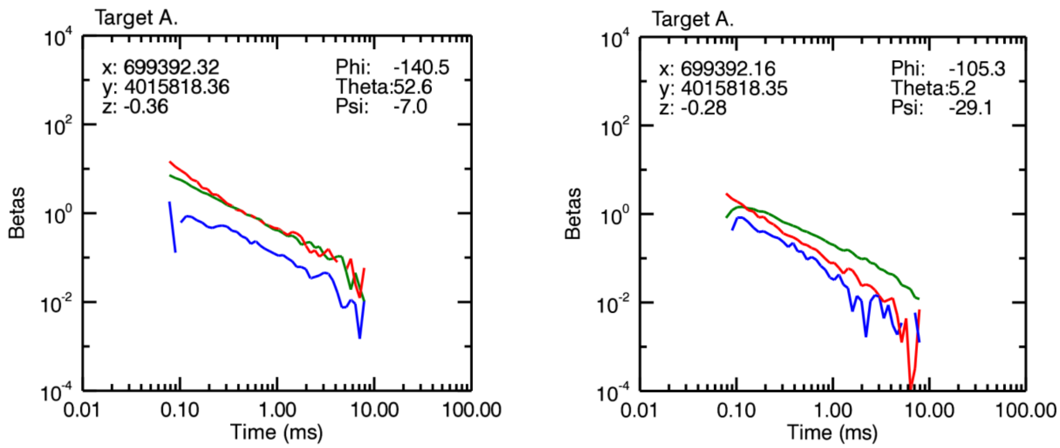


Figure 6-51. Polarization plots of anomaly 2504 after removing Rx3Y data using the single-dipole solver (left) and multi-dipole solver (right).

The only other false negative was anomaly 1346. The ground truth indicates this is a 37mm buried at a depth of 30cm. This anomaly was surveyed with both the Geometrics and Sky MM systems. Upon closer inspection of the data, the Geometrics system had problems with receiver coil 3Y. Figure 6-52 and Figure 6-53 shows polarization plots before and after removing the data for Rx3Y. The plan view in Figure 6-53 shows two clusters of green diamonds which indicates the existence of two sources. One of the parameters of the multi-dipole solver controls the separation between sources. The setting used was adequate for almost all the anomalies but should have been smaller for this anomaly as the final solution was a single dipole with a location between the two clusters. Figure 6-54 shows the fit results if we reduce the separation parameter. Now, the solver produces two sources centered on each of the clusters. The target labeled "A" matches best to a 37mm but the 3-criteria match was low because the polarizations are noisy and the 3rd polarization does not match the 2nd. The 3rd polarization is the first to give erroneous results as the signal to noise decreases as is the case for this anomaly. The 2-criteria and 1-criteria matches were 0.65 and 0.92, respectively.

Figure 6-55 shows an overview of Sky MM data and a close up section. The oval highlights an area with elevated responses which indicates an improper background was removed. Anomaly 1346 was located in this area and the background problem may have affected the polarizations enough to cause the failure. The 2-criteria library metric for the single solver was 0.8029 which was just below the cutoff of 0.81 so a small improvement in the polarizations would have been sufficient to classify this target as a "Dig".

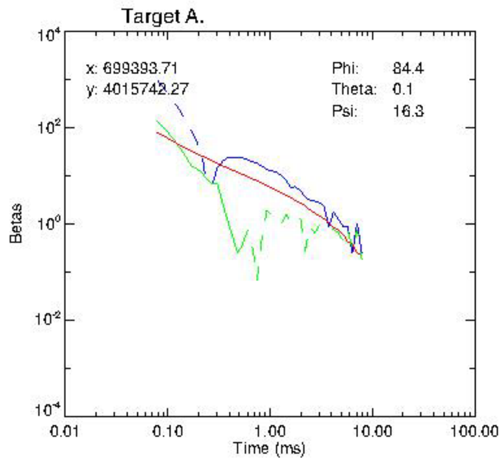


Figure 6-52. Polarization plot of anomaly 1346 using the Geometrics MM and the multi-dipole solver.

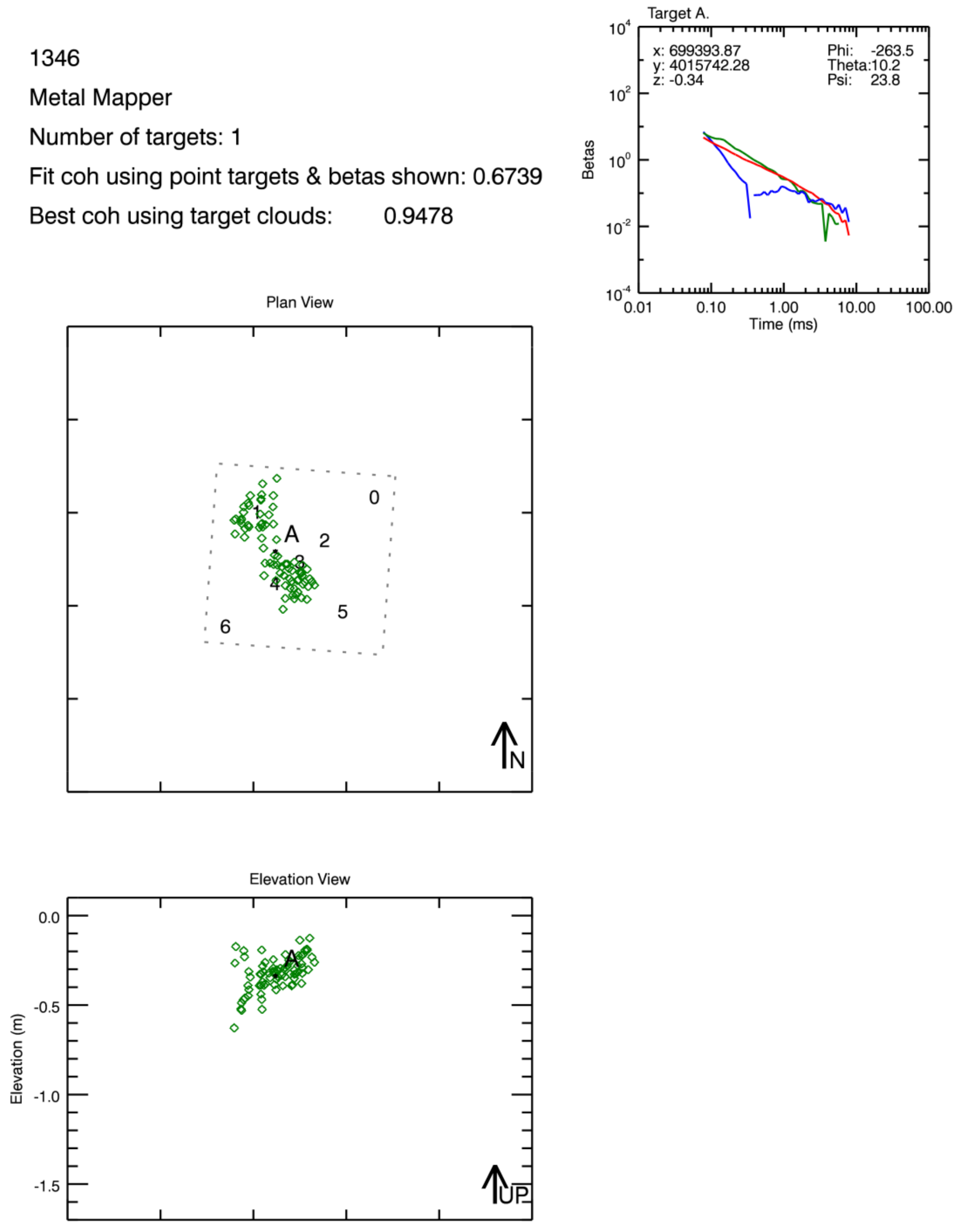


Figure 6-53. Fit results plot of Geometrics' MM anomaly 1346 after removing Rx3Y. The green diamonds represent weights assigned to hypothetical sources by the algorithm.

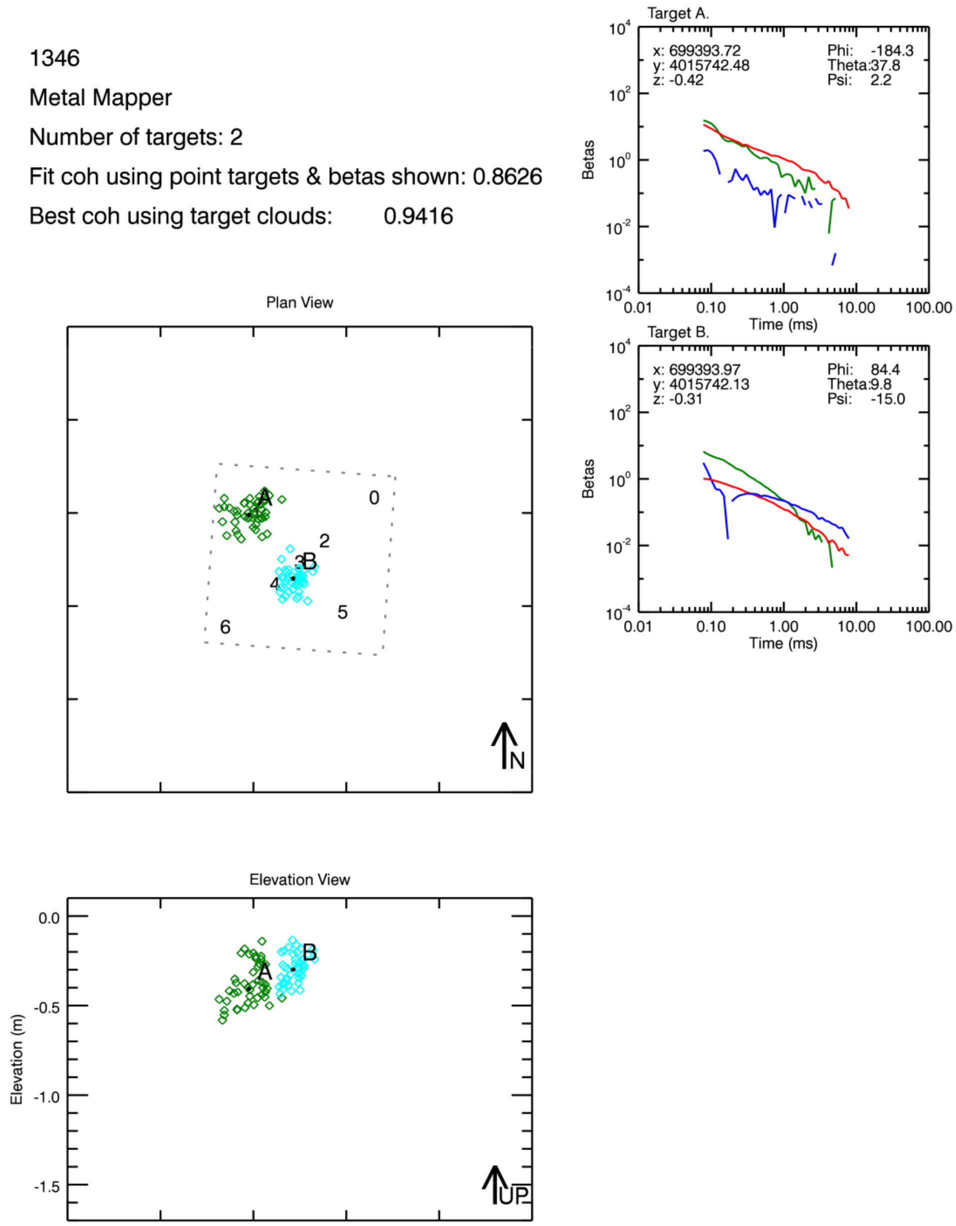


Figure 6-54. Fit results plot of Geometrics' MM anomaly 1346 after removing Rx3Y and adjusting the multi-dipole solver parameters.

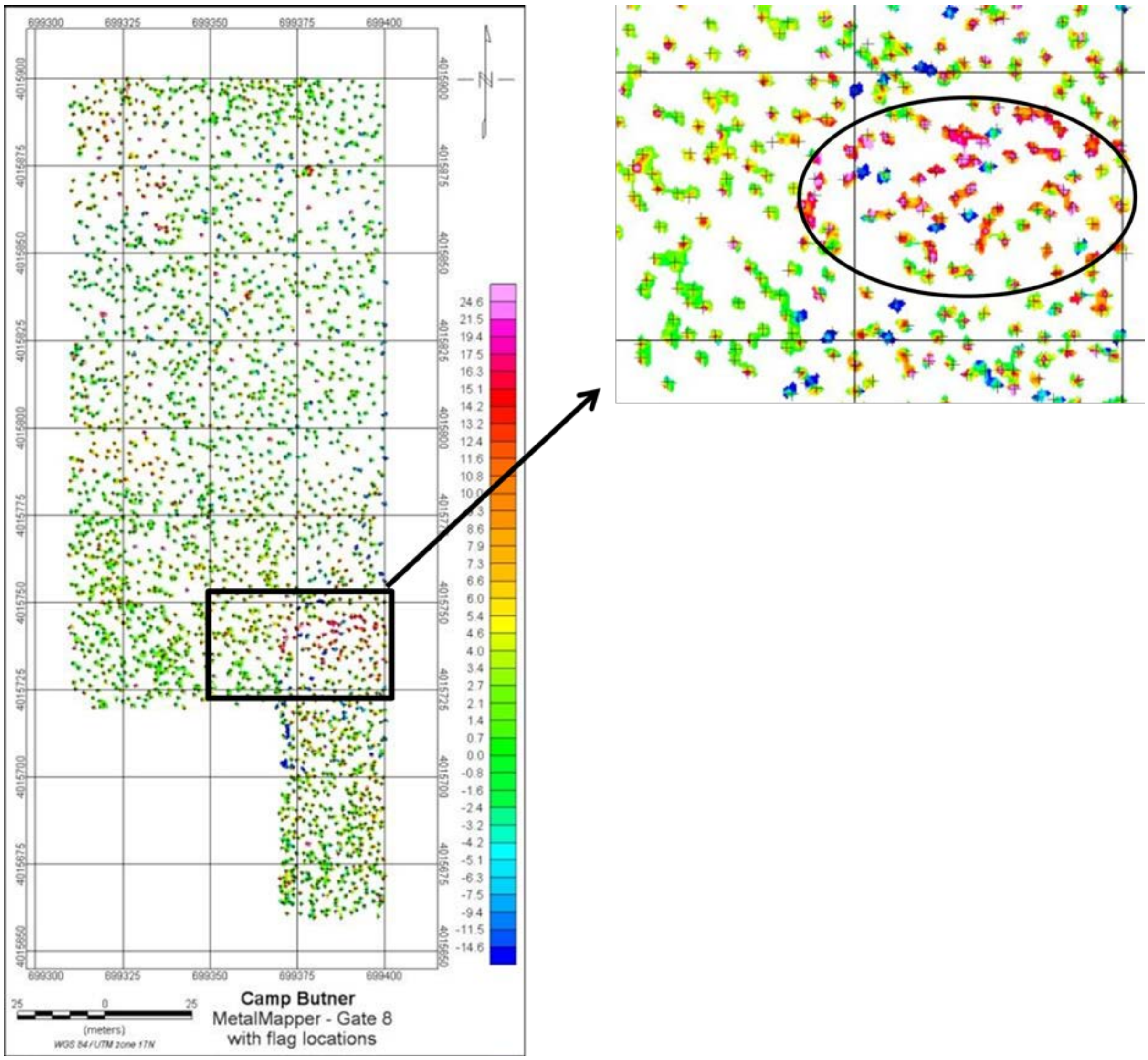


Figure 6-55. Overview and zoomed section of the Sky MetalMapper data. The oval highlights an area with elevated response that contains anomaly 1346.

6.5.7 METAL MAPPER - IDL

Data from the Metal Mapper are shown in Figure 6-56. The crosses identify anomalies that were excavated and scored by the ESTCP Program Office.

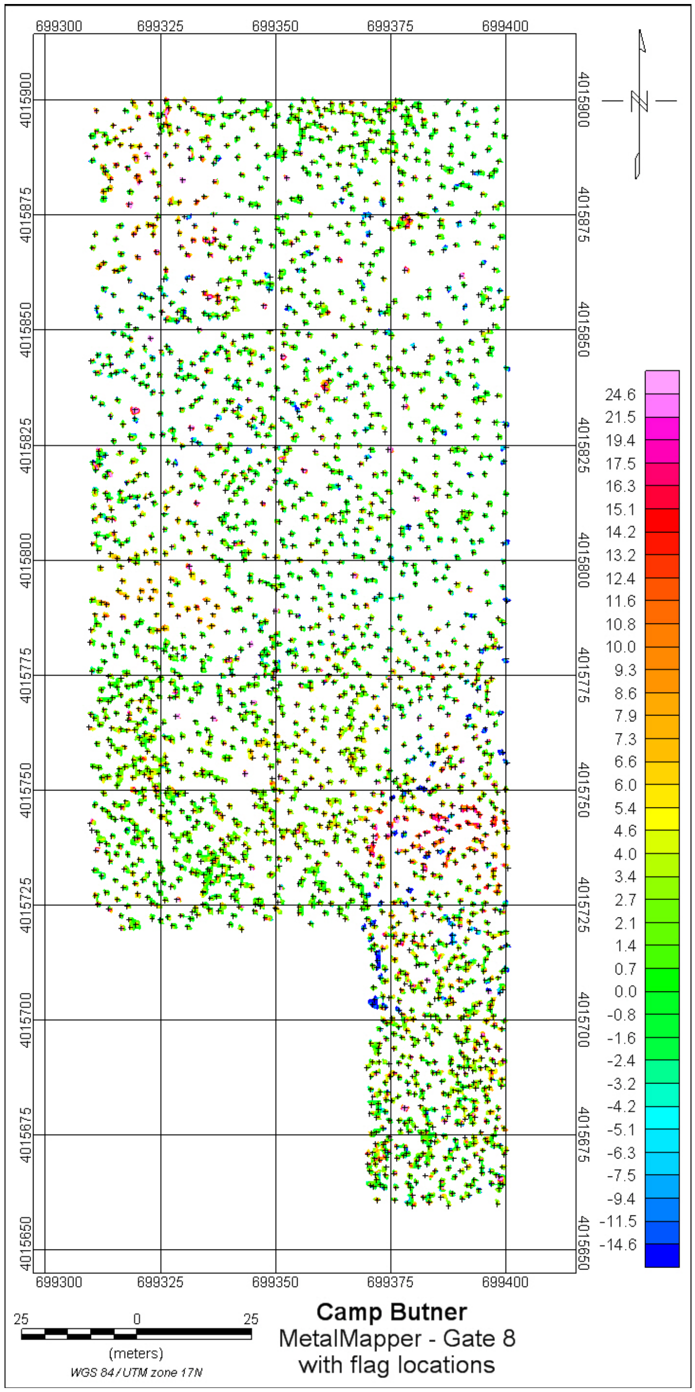


Figure 6-56. Metal Mapper data with anomaly locations.

Performance Scores from IDA

Scoring performances for the NOSLN, 2 and 3 criteria MM analysis using IDL are reported in Table 6-20 to Table 6-23. Their respective ROC charts are shown in Figure 6-57 to Figure 6-60, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 13, 7, 0 and 0 false negatives for the NOSLN 1st pass, NOSLN 2nd pass, 2-criteria and 3-criteria analysis methods, respectively. All the false negatives and all troublesome Category 2 anomalies were 37mm. All the lists used the same target features but with a different library match metric and classification rules.

Table 6-20 Test Set Summary: IDL - MM – NOSLN 1st pass

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	32	1078	11	13	1134
2	12	826	25	12	875
3	1	87	3	146	237
4	0	43	1	0	44
Training	0	0	0	0	0
TOTAL	**45**	**2034**	**40**	**171**	**2290**

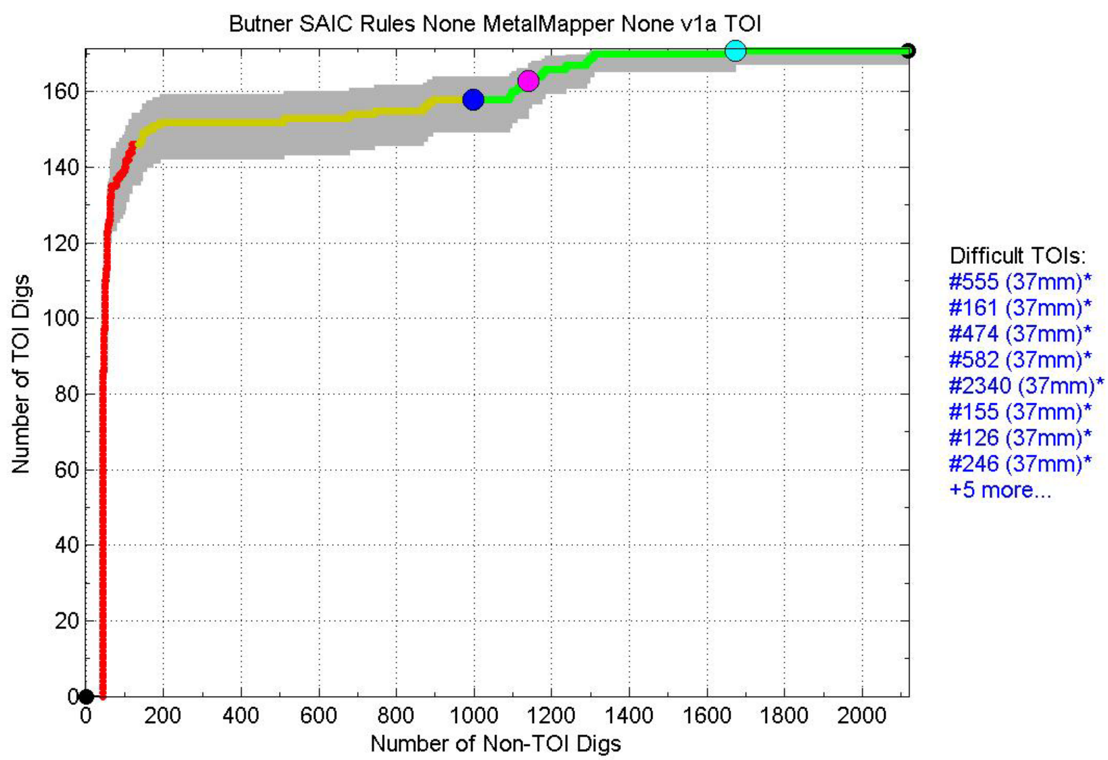


Figure 6-57. IDL - MM – NOSLN 1st pass ROC chart.

Table 6-21 Test Set Summary: IDL - MM – NOSLN 2nd pass

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	31	1068	11	7	1117
2	7	691	24	6	728
3	1	94	1	9	105
4	0	43	1	0	44
Training	6	138	3	149	296
TOTAL	**45**	**2034**	**40**	**171**	**2290**

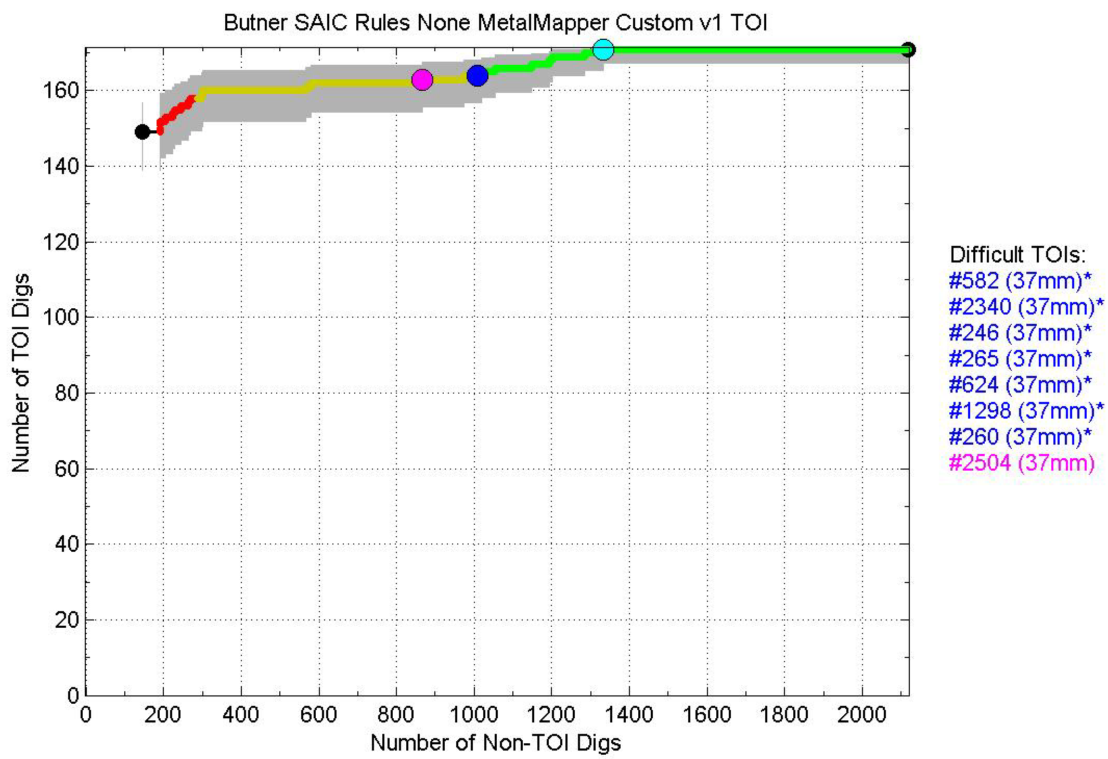


Figure 6-58. IDL - MM – NOSLN 2nd pass ROC chart.

Table 6-22 Test Set Summary: IDL - MM – 2-criteria

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	7	228	2	0	237
2	18	1265	27	5	1315
3	14	407	7	156	584
4	0	43	1	0	44
Training	6	91	3	10	110
TOTAL	**45**	**2034**	**40**	**171**	**2290**

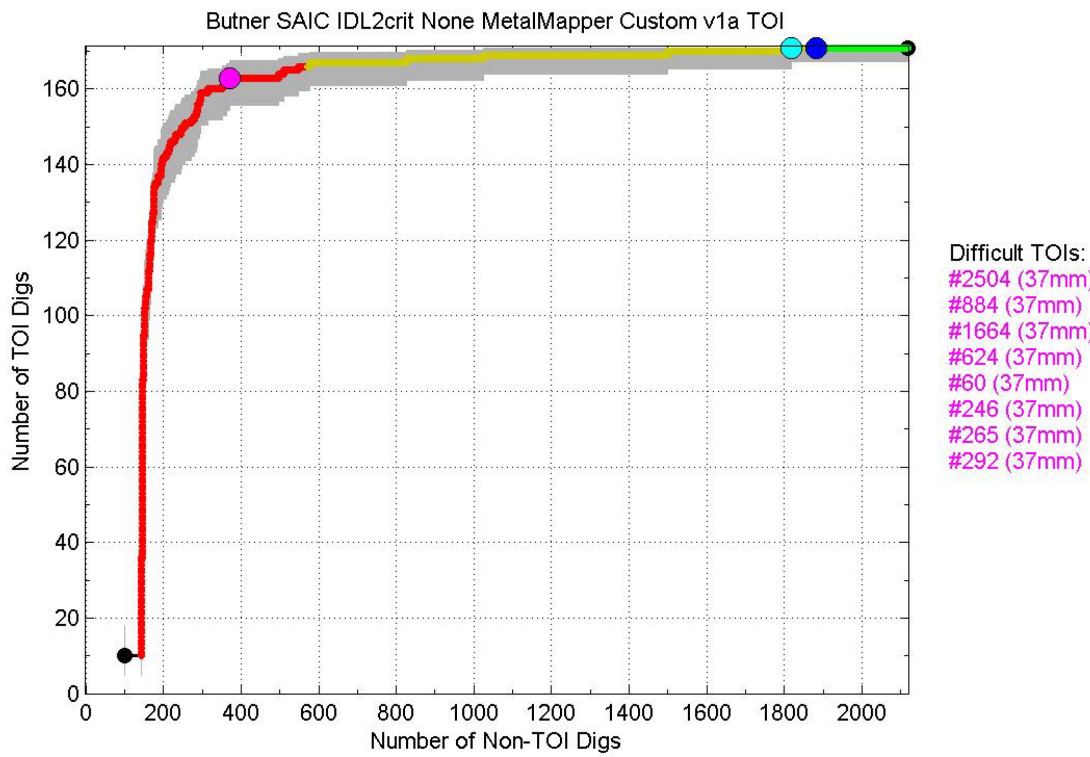


Figure 6-59. IDL - MM 2-criteria ROC chart.

Table 6-23 Test Set Summary: IDL - MM – 3-criteria

Category	Cultural	Munition Debris	No Contact	UXO	TOTAL
1	10	283	2	0	295
2	24	1486	32	14	1556
3	5	126	4	148	283
4	0	43	1	0	44
Training	6	96	1	9	112
TOTAL	**45**	**2034**	**40**	**171**	**2290**

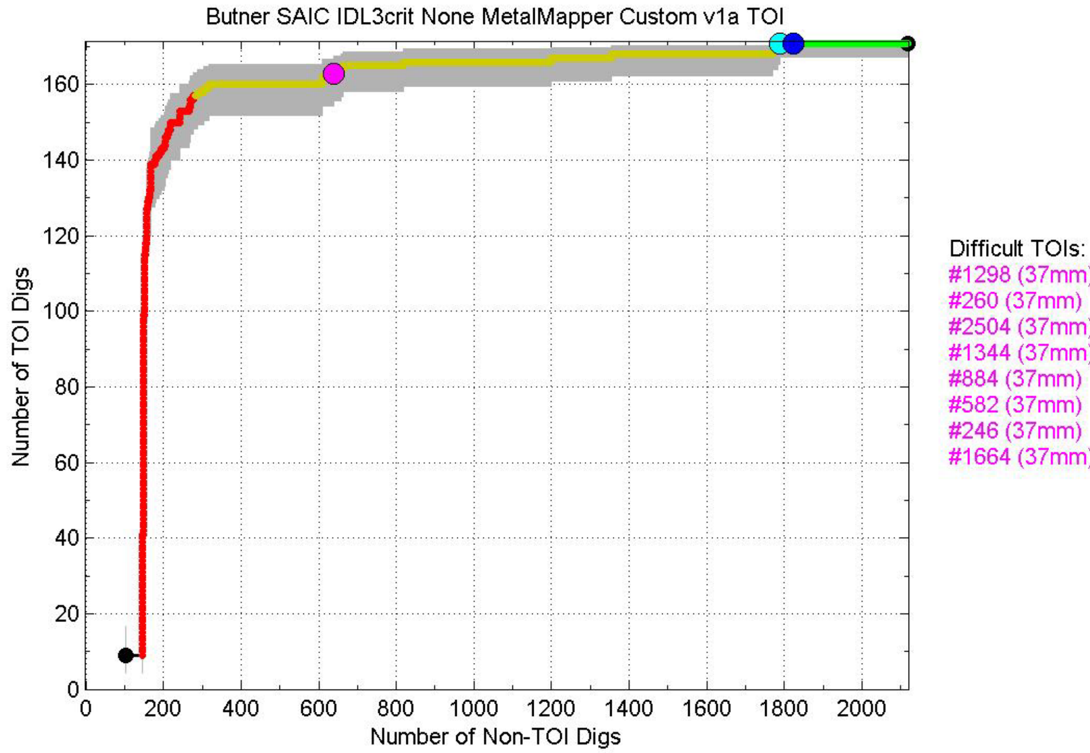


Figure 6-60. IDL - MM 3-criteria ROC chart.

Characterization Plots

Figure 6-61 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets. The mean error for all TOI with an isolated or slightly overlapping signal was 0.12m with a standard deviation of .08m. If non-TOI are added to the population the mean error increases to 0.18m with a standard deviation of 0.20m.

Figure 6-62 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets. The mean error for all TOI with an isolated or slightly overlapping signal was 0.055m with a standard deviation of .06m. If non-TOI are added to the population the mean error was 0.016m with a standard deviation of 0.09m.

Figure 6-63 shows the inverted polarizabilities for all category 1, 2 and 3 targets. Although not as tight as the TEMTADS, there still is good clustering of the TOI which allows the use of shape to characterize the targets. Table 6-24 tabulates the statistics of the polarizations for the different TOI after removing those anomalies with large fit errors.

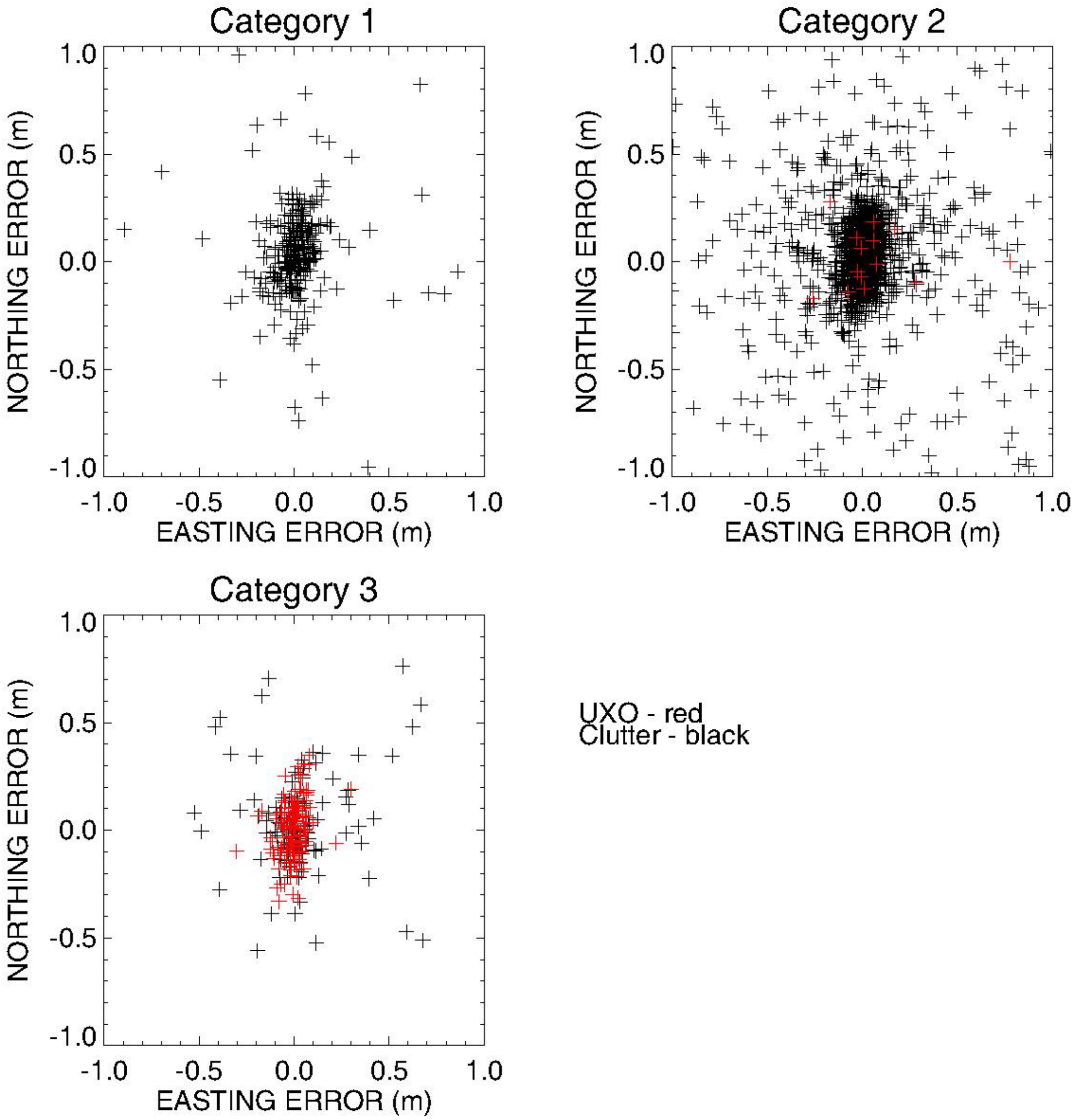


Figure 6-61. Differences between fitted and measured XY locations; Metal Mapper 2-criteria analysis

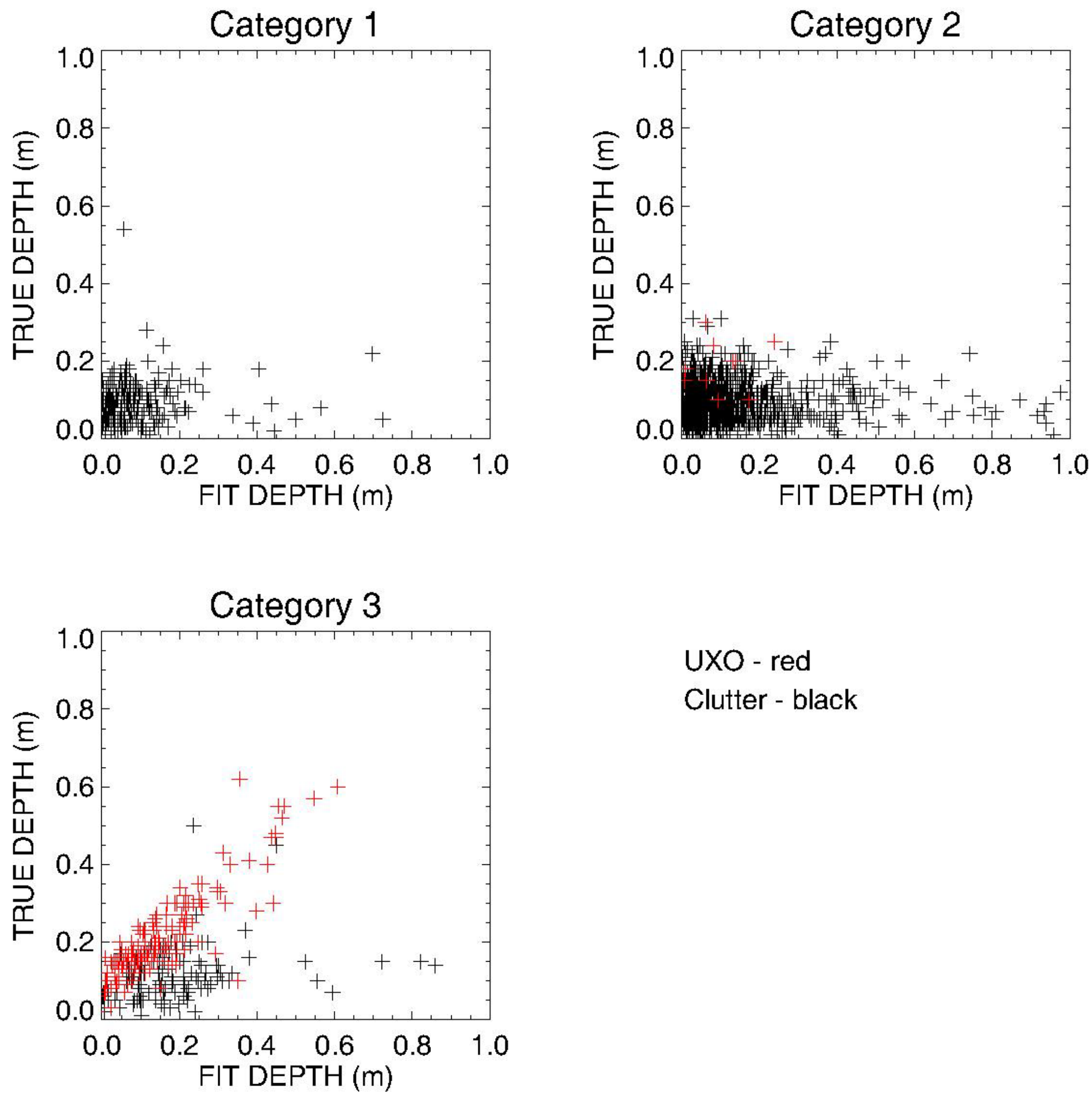


Figure 6-62. Fitted versus measured depth of burial; Metal Mapper 2-criteria analysis

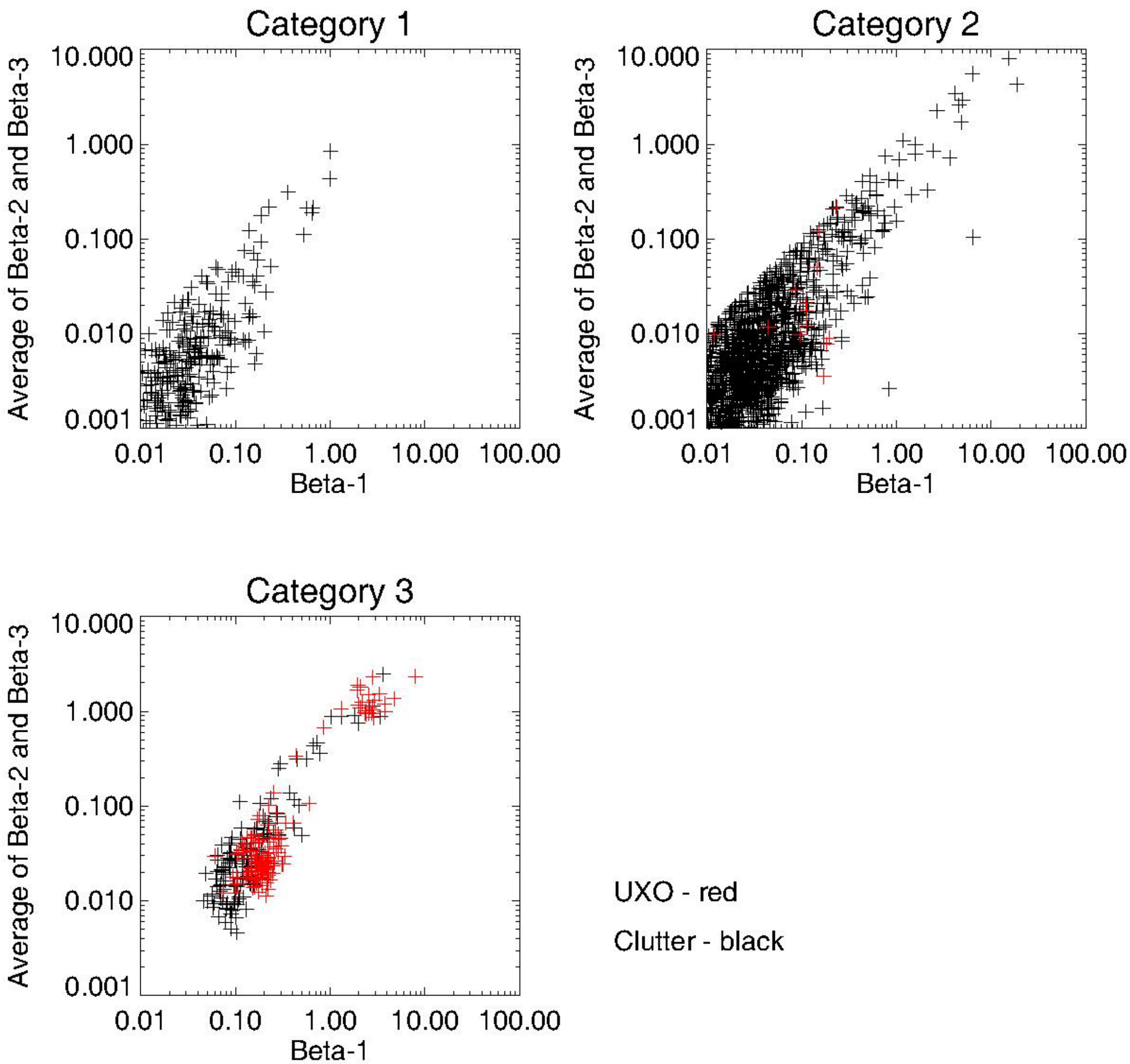


Figure 6-63. Beta1 versus the average of Beta 2 and Beta 3; IDL-MM 2-criteria analysis.

Table 6-24 Statistics of betas for the three main TOI, IDL – MM

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
37mm	104	0.038	0.005	0.201	0.090	0.029	0.020	0.015	0.019
Fuze	22	0.038	0.003	0.152	0.034	0.048	0.018	0.037	0.011
105mm	26	0.105	0.011	2.775	1.306	1.278	0.420	0.962	0.314

Failure Analyses

In Table 6-25, we give the Anomaly IDs of our false negatives for all the IDL based MM only analysis methods as well as problematic Category 2 anomalies that forced a late "Dig"/"Do not Dig" cutoff. Both the 3 and 2 criteria methods did not have a false negative but they also did not declare many targets as Category 1 or "Do not dig". The vast majority of the problem anomalies came from decisions based on the Geometrics MM.

Table 6-25 Metal Mapper's False Negatives

Anomaly ID	NOSLN 1st pass	NOSLN 2nd pass	2 criteria Cat 2	3 criteria Cat 2	Sensor
99	X				Geometrics
126	X				Geometrics
155	X				Geometrics
161	X				Geometrics
246	X	X		X	Geometrics
260		X		X	Geometrics
265	X	X			Geometrics
474	X				Geometrics
555	X				Sky
582	X	X		X	Geometrics
624	X	X	X		Geometrics
884			X	X	Geometrics
1298		X		X	Geometrics
1344				X	Geometrics
1346	X				Sky/Geometrics
1664			X	X	Geometrics
2017	X				Geometrics
2340	X	X			Geometrics
2504			X	X	Geometrics

There were two problematic anomalies that were surveyed by the Sky MM system. The failure analysis of anomaly 1346 was discussed in Section 6.5.6. The other anomaly was 555. This anomaly could be resolved by using the inversion result from the tensor solution instead of the eigenvalues solution. As discussed in Section 6.5.3 the eigenvalues solution uses a constant orientation as a function of time gate during the data inversion whereas the tensor solution allows the orientation to vary across the different time gate. The tensor solution produces a very good match to a 37mm.

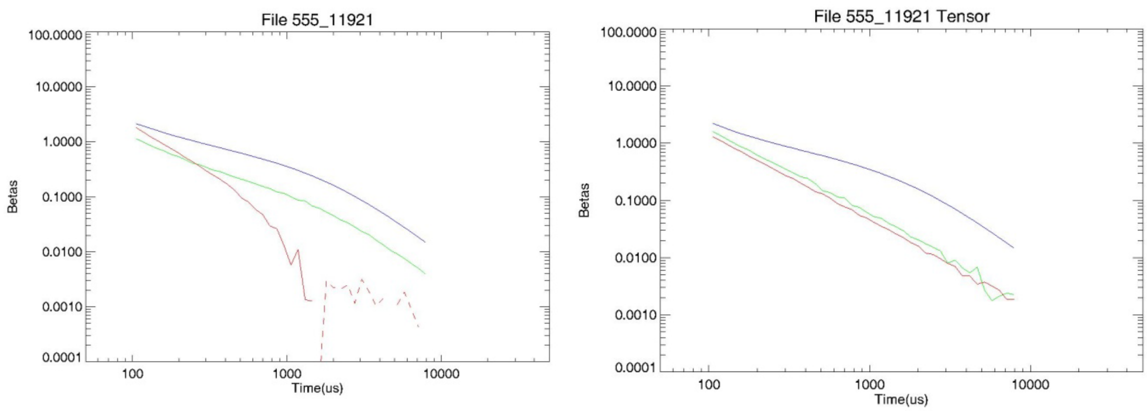


Figure 6-64. Polarization plots for Anomaly 555 using data from the Sky MM. The plot on the left shows the eigenvalue solution while the one the right is the tensor solution.

The remaining problematic anomalies all used data from the Geometrics MM. There a few different issues that caused the misclassifications but the primary reason was corrupt data from the Y-axis component of receiver coil 2 which we refer to as Rx2Y. The IDL analysis team was aware of problems with Rx3Y which was discussed in the previous section but was not aware of problems with Rx2Y. During our retrospective analysis of the failures a bug in the way bad data was being removed from the inversion routines was also discovered. We do a non-linear fit on X, Y, Z (and sometimes the 3 angles), but embedded in this is a linear solution for the betas. The problem was that the zero weights for the problem coils were not passed to the linear regression. Therefore, the linear solution for the betas included the bad coils, even though the non-linear regression did not. Figure 6-65 shows the measured decay QC plot for anomaly 260 with the bad data from Rx2Y circled. Figure 6-66 shows the polarization plot of the originally submitted data on the left and after removing the data from Rx2Y on the right. The original polarizations consisted of one negative polarization and two positive polarizations that did not match to a 37mm. After removing the Rx2Y data the three polarizations were positive but they also did not match to a 37mm. Figure 6-67 shows the polarizations after removing Rx2Y and passing the data through the updated inversion routines which properly handle the removed coils. These polarizations produced a very good match to a 37mm. Using the same procedures, similar results were obtained for anomalies 99, 126, 155, 161, 246, 265, 474, 624, 1298, 1344 and 1664.

Anomalies 582, 884 and 2017 are resolved by using the tensor solution in addition to above procedures. Anomaly 2304 did not have problems with Rx2Y but did have problems with Rx3Y as discussed in the previous section. The cause of failure for this anomaly was the bug in the inversion algorithm that caused the bad data from Rx3Y to be used.

The remaining failure was anomaly 2504. This anomaly was discussed in Section 6.5.6 with the cause of the misclassification being low SNR.

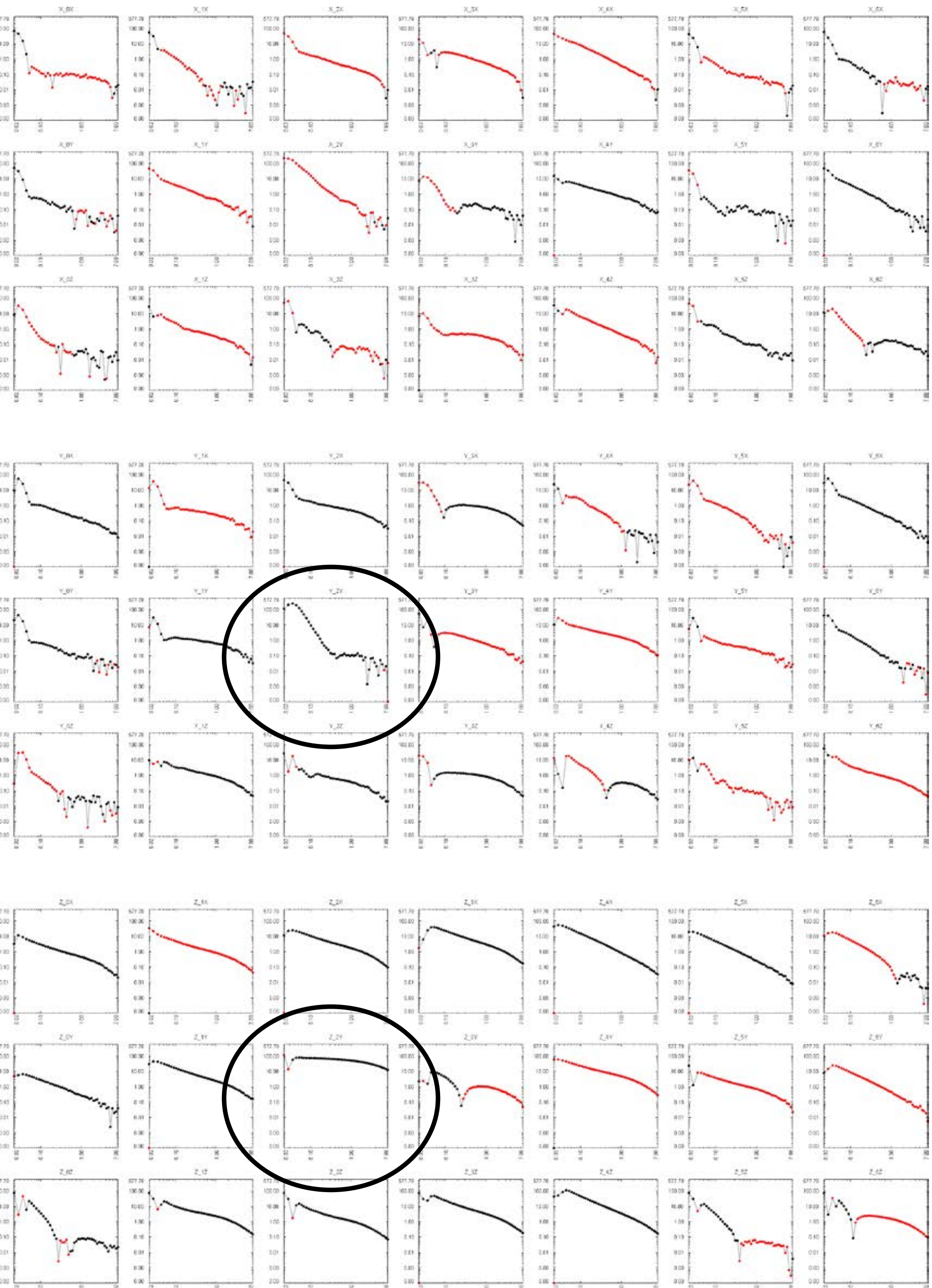
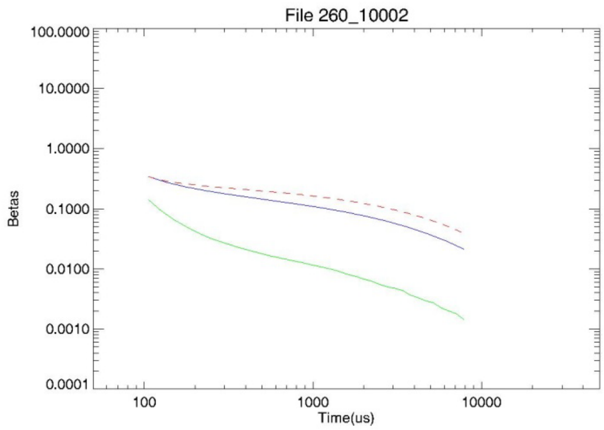


Figure 6-65. Measured decay QC plot of the MM data for anomaly 260 with the problematic Rx2Y data circled.



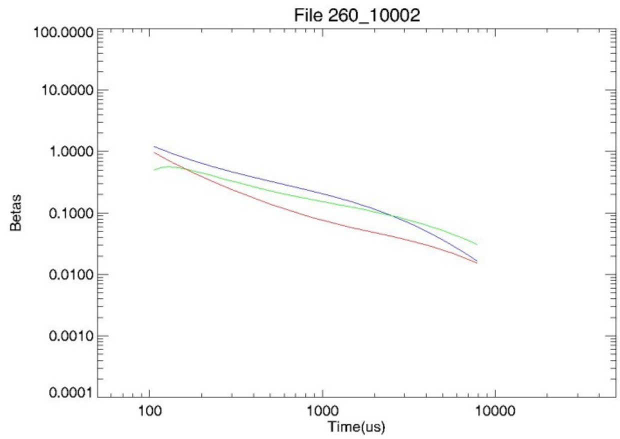


Figure 6-66. Polarization plots for Anomaly 260 using data from the Geometrics MM. The plot on the left shows polarizations as submitted while the one the right shows the polarizations after removing the bad data from Rx2Y.

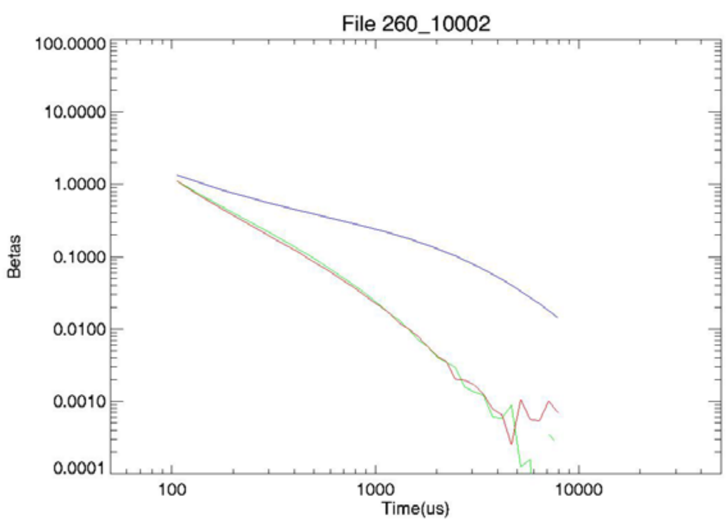


Figure 6-67. Polarization plot for Anomaly 260 with the bad data from Rx2Y removed and using the updated inversion routines.

6.5.8 UXANALYZE – METAL MAPPER - NAEVA

NAEVA used UX-Analyze to characterize and classify the Metal Mapper data. They submitted three dig sheets using the same target features but with slightly different classification rules. All the dig sheets used the 3-criteria library matching metric found in UX-Analyze and the standard training set. The first dig sheet only used the 3-criteria metric. The second dig sheet used the 3-criteria and the 2-criteria metrics to classify the anomalies. The last dig sheet used the 3-criteria in conjunction with the 1-criteria metric. Details on NAEVA's analysis methods can be found in Appendix B.

Performance Scores from IDA

Scoring performances for NAEVA's UX-Analyze MM analyses are shown as a ROC charts in Figure 6-68 to Figure 6-70, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for these analyses, there was one false negative using the 3-criteria method. Anomaly 1346 was classified as high confidence clutter.

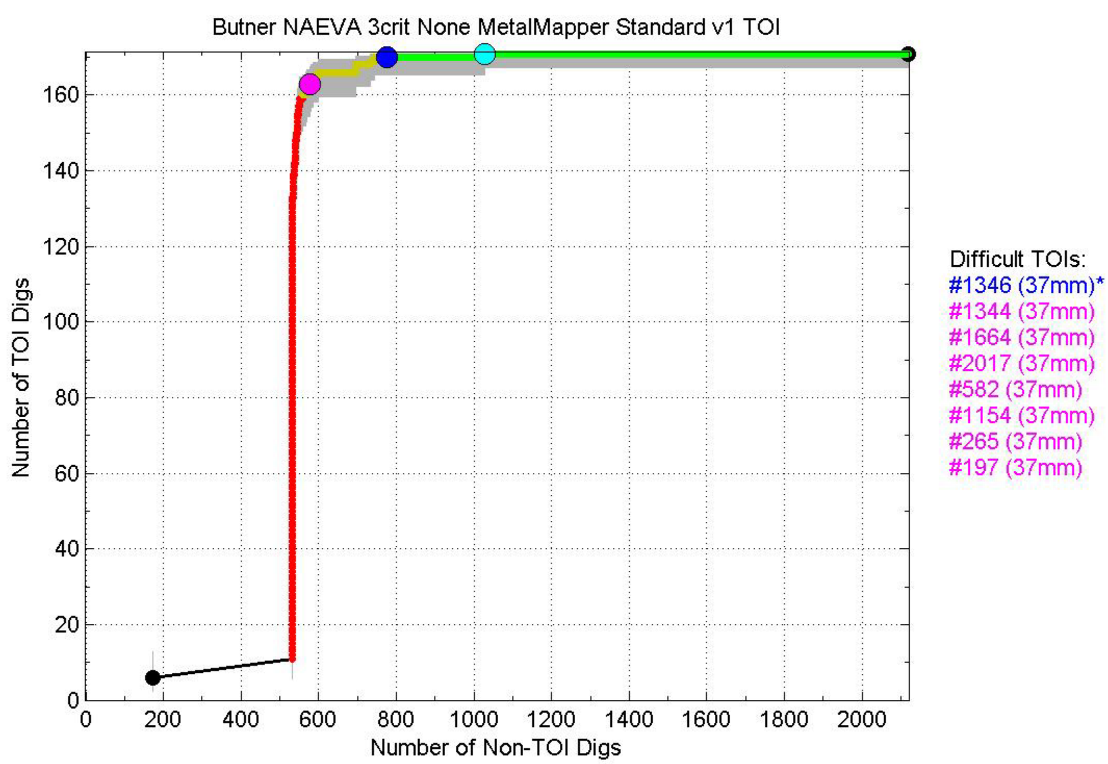


Figure 6-68. NAEVA's UX-Analyze Metal Mapper 3-criteria ROC chart.

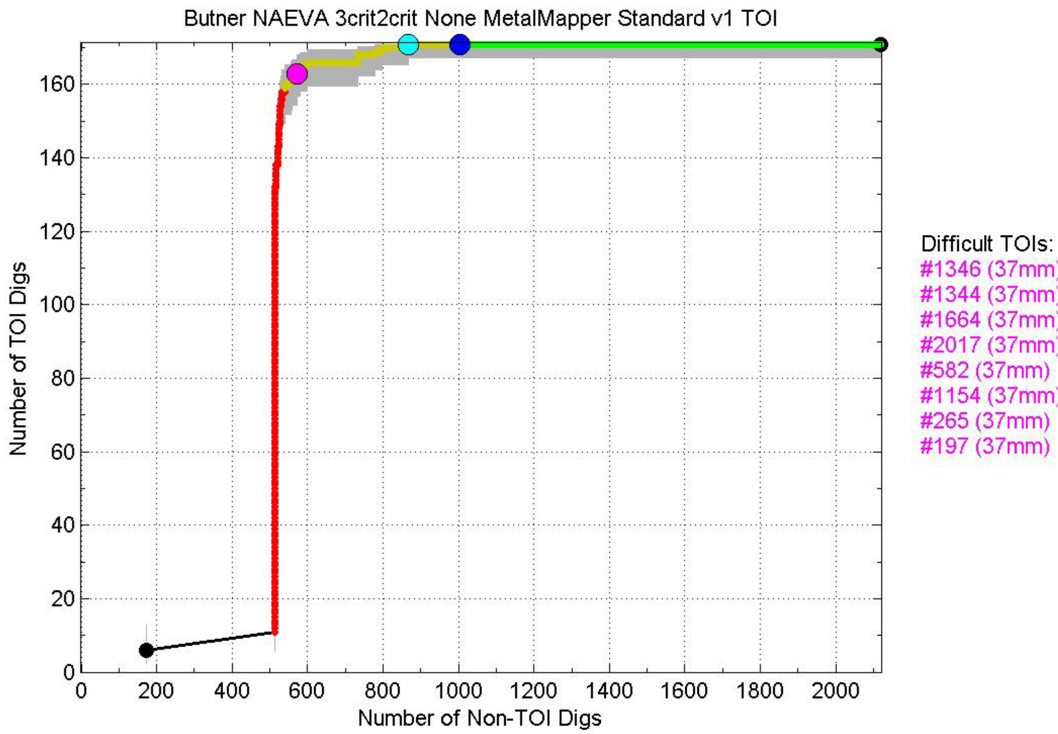


Figure 6-69. NAEVA's UX-Analyze Metal Mapper 3-criteria/ 2-criteria ROC chart.

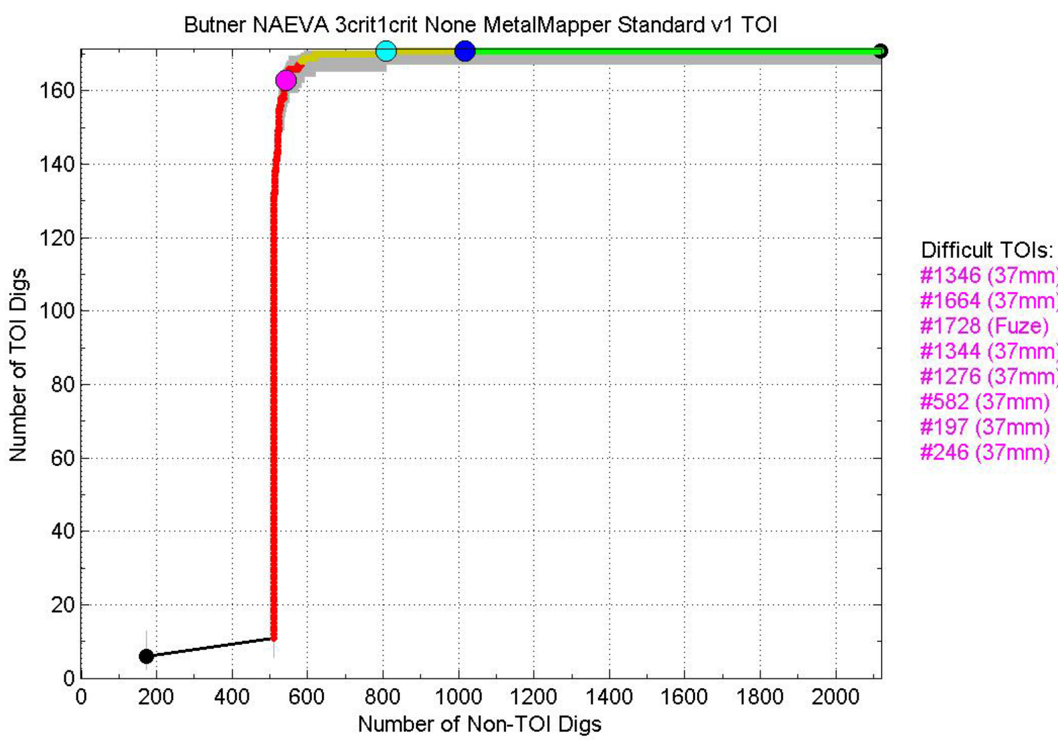


Figure 6-70. NAEVA's UX-Analyze Metal Mapper 3-criteria/1-criteria ROC chart.

6.5.9 UXANALYZE – EM61-MK2 with METAL MAPPER - PARSONS

The EM61-MK2 data was used as a first pass to try to discriminate between TOI and clutter and removed approximately 500 targets. The remaining targets were classified using MM data. Specific details on the different processing methods and failure analysis can be found in Parsons' report which is included in this report as Appendix C.

Performance Scores from IDA

Scoring performance for Parsons' UX-Analyze EM61-MK2/MM analysis is shown as a ROC chart in Figure 6-71, where we plot the Number of TOI Digs versus the Number of Non-TOI Digs.

Using the thresholds adopted for this analysis, there were 5 false negatives. Anomaly 404 was classified as clutter using the EM61-MK2 data. The other anomalies (272, 884, 1154 and 1344) were classified as high confidence clutter using the MM data.

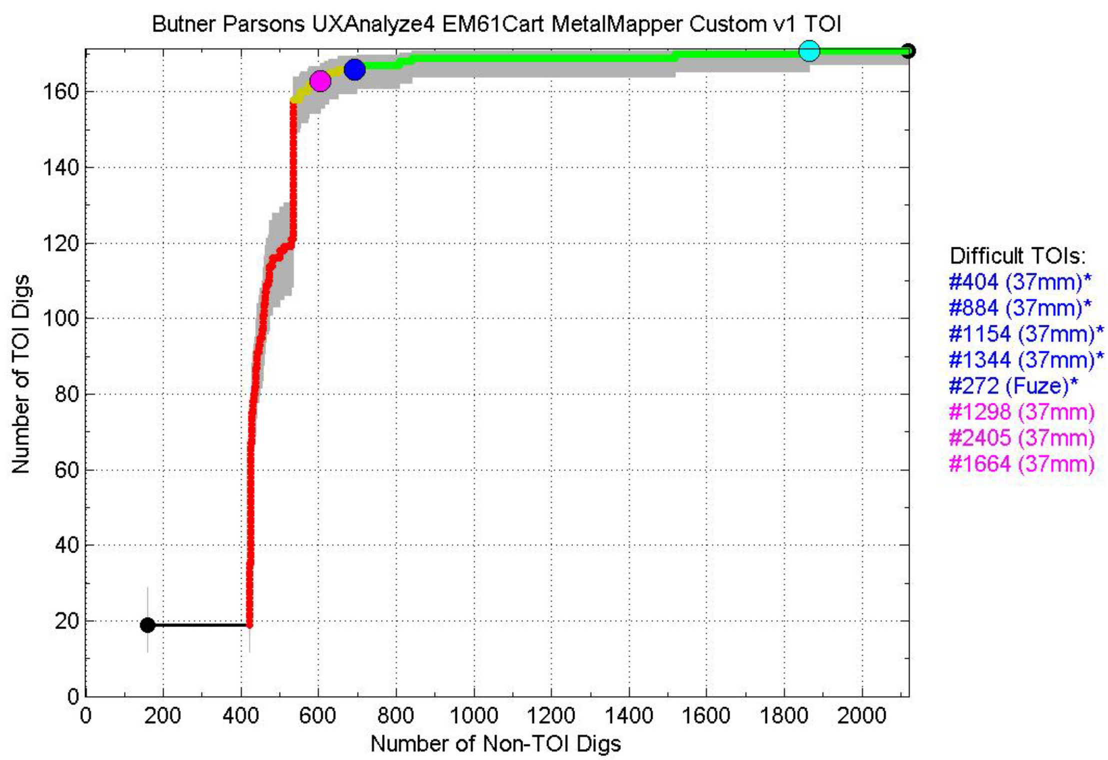


Figure 6-71. Parsons' UX-Analyze EM61-MK2 Cart with Metal Mapper ROC chart.

6.6 DISCUSSION

In general, the analysis methods that used the dynamically collected EM61-MK2 data either alone or as a pre-screener did not perform as well at this site. At past sites the decay ratios and inverted target size were very useful features to help discriminate between some of the clutter and the TOI. Unfortunately, at this site the EM61 data were only able remove about 15% of the clutter items and still produced a false negative. Using the EM61 as a pre-screener reduced the number of cued targets by about 15-20% but also resulted in one or two false negatives. Overall, the EM61-MK2 data were not very useful for discrimination at this site because the target features for the clutter and the TOI had significant overlap.

In contrast, the cued fixed array systems consisting of the TEMTADS and Metal Mapper produced polarizations that were accurate enough to discriminate between TOI and non-TOI on the basis of shape. The ROC curves are similar for the TEMTADS and Metal Mapper although the TEMTADS' curve tends to start more vertical than the Metal Mapper's. Both of the cued sensors had ROC curves that were much better than those for the EM61 sensor. For most of the analysis methods using only the cued data, there are few false positives from the beginning of the curve until you reach about 95% of TOI recovered. The remaining 5% of TOI require a large number of excavations for most of the dig lists using the thresholds defined and typically resulted in single digit false negatives. All the submittals were ranked by decision rules with no "human veto" or interactive classification.

The false negatives can be divided into two main types. The first is comprised of data problems. These occurred for both the TEMTADS and the Metal Mapper sensors. For the TEMTADS sensor it was discovered that the polarity of coil 5 was reversed. The degree of degradation of the fitted results was a function of the proximity of the source to this coil. Luckily, coil 5 was on the perimeter of the sensor and the overall effect was minimal for most anomalies but when the source was near the problem coil there were adverse consequences. The Geometrics MM system had intermittent problems with the Y axis coil for receivers 2 and 3. Because this was an intermittent problem it did not affect the majority of the anomalies but when the problematic data were used in the inversion algorithms the results were affected. In order to overcome these data problems better communication and QC of the data was needed. In the case of the MM problem it was known during data collection that intermittent problems occurred for Rx2Y and Rx3Y but the analysis teams were aware of only one or the other. Although not informed of the data problems the analysis teams could have discovered the problem data with a more stringent inspection of the inverted results.

The second category of false negatives is multiple targets. SAIC has made good progress on developing a reliable multi-dipole solver under SERDP project MR-1662. As improvements are made to the multi-dipole solver they are easily transitioned to UX-Analyze. The IDL based analysis dig sheets only used single-dipole solvers and a few of their false negatives would have been caught if a multi-dipole solver had been used. The multi-dipole solver also reduces the manual work needed to carve out portions of the data collected that are affected by a secondary source prior to passing the data to the single-dipole solvers. This is also only done by visual

inspection and is prone to error especially if one source is shallow and the other is directly below it and deep. Although using the multi-dipole solver has been very useful it still needs some improvements and more testing. There are a number of parameters that require setting that control the sensitivity of the solver and even though our default values produced good results for virtually all the anomalies there were a handful that would have benefitted with slightly different parameters. Therefore, we found during this demonstration that using a combination of the results from the multi-dipole solver and the single-dipole solver and keeping the one most like a TOI was the safest method of analyzing the data. However, after accounting for the data problems with coil 5 of the TEMTADS sensor we found that using only the multi-dipole solver was sufficient to capture all the TOI. The results of this test will be discussed later in this section.

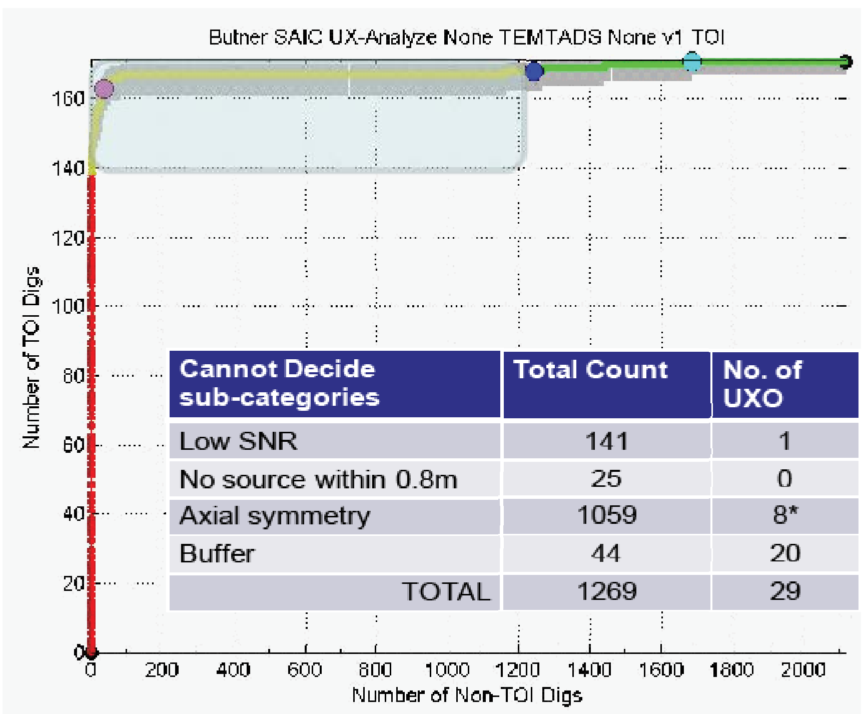
A common theme for many of the ROC curves was a large number of Category 2 anomalies with the most common subcategory being anomalies that met our axial symmetry metric. The library used contained only expected targets at the site. Therefore, the axial symmetry metric was inserted to account for objects that were not in our library but could potentially be a UXO. Figure 6-72 shows the ROC curve for the UX-Analyze TEMTADS data analysis with a breakdown of the number and type of “Cannot Decide” anomalies. As the figure shows most of the “Cannot Decide” anomalies were due to the axial symmetry metric and out of the 1059 only 8 were TOI. All 8 of the TOI were also declared at the beginning of the “Cannot Decide” anomalies. If we modified our UXO/clutter threshold and did not hedge for unexpected munition types by eliminating the axial symmetry metric we would have reduced the number of unnecessary digs by 951 with the same probability of classification.

Eliminating the axial symmetry metric did not affect the probability of classification at this site because there were not any unexpected UXO types. However, that is not the case for many sites so our decision rules still need to allow for unexpected UXO. An alternate method was tested that compared the primary polarization to a library containing all types of UXO and not only the expected ones. This “1-criteria” metric uses the primary polarization which gives an estimate of size. This should be the most stable metric because as SNR decreases we have found the secondary and tertiary polarizations are the first to degrade. In addition to the 1-criteria metric a few other changes to the decision rules were made. The main difference was the use of three library metrics instead of just the 2-criteria metric. The 3 and 2 criteria metrics were calculated using a library containing only the expected TOI while the 1-criteria metric used a library containing all UXO. The 3 and 2 criteria metrics were only used to extract Category 3 targets with all targets having a high 3-criteria metric being classified as high confidence UXO and ranked the highest followed by targets with a high 2-criteria metric. The last Category 3 subcategory was targets with a high 1-criteria metric. Virtually the same decision rules were used to define the Category 1 and 2 targets except the 1-criteria metric was used to make the decision and not the 2 or 3 criteria metric that was used in our original submissions.

Figure 6-73 shows the ROC curve of the TEMTADS data analysis using the new decision rules, fixing the polarity of coil 5 and using both the single and multi-dipole solver. Comparing this ROC curve to the ROC curve in Figure 6-72 shows a couple of significant improvements. First,

the number of “Cannot Decide” anomalies has been greatly reduced. Second, a larger percentage (87%) of the clutter items was declared “Do not Dig” without incurring a false negative. In fact all the TOI were correctly classified as Category 3 anomalies. Figure 6-74 shows the ROC curve if we only use the multi-dipole solver. Here we see even better performance as 90% of the clutter items were declared “Do not Dig” without incurring a false negative. By fixing the polarity problem associated with coil 5 all the problematic anomalies were able to be resolved using only the multi-dipole solver.

Figure 6-75 shows the ROC curve of the MM data analysis using the new decision rules, removing data from Rx2Y and Rx3Y and using both the single and multi-dipole solver. Comparing this ROC curve to previous ones also shows significant improvement. Similar to the updated TEMTADS analysis, the number of “Cannot Decide” anomalies has been reduced and 59% of the clutter items were declared “Do not Dig” without incurring a false negative. Although the results of the MM analysis are very good, overall it did not perform as well as the TEMTADS data analysis at this site using the described methodology.



Cannot Decide sub-categories	Total Count	No. of UXO
Low SNR	141	1
No source within 0.8m	25	0
Axial symmetry	1059	8*
Buffer	44	20
TOTAL	1269	29

Figure 6-72. ROC curve of UX-Analyze TEMTADS analysis and associated Cannot Decide subcategories.

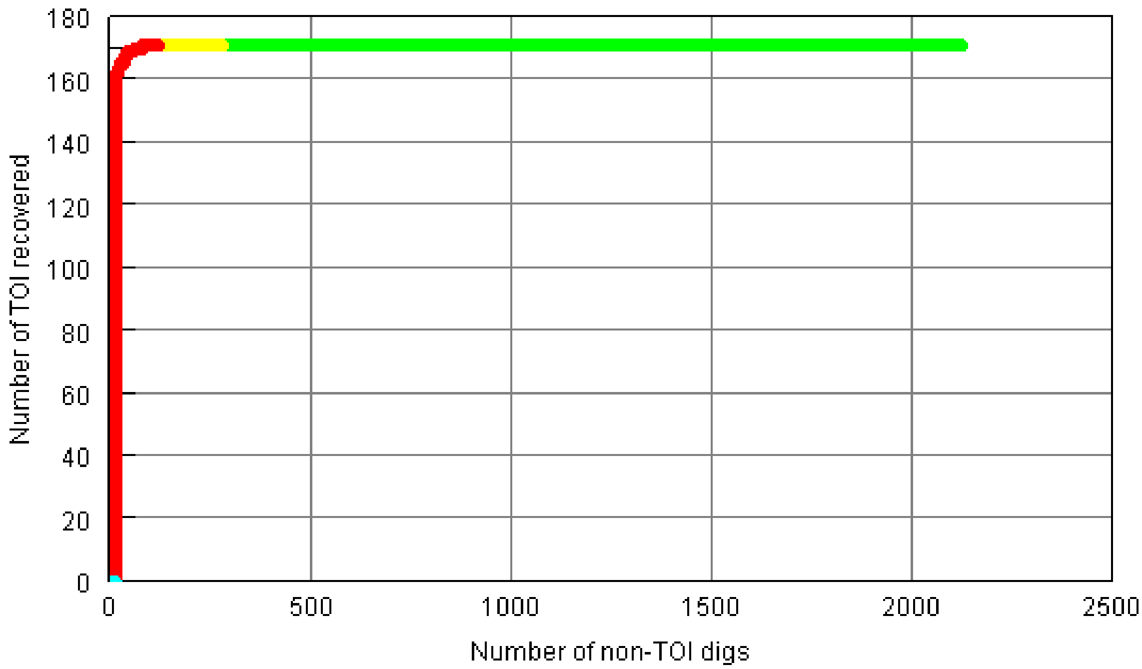


Figure 6-73. The ROC curve of the TEMTADS data using both the single and multi-dipole solvers, new decision rules and fixing the polarity of coil 5.

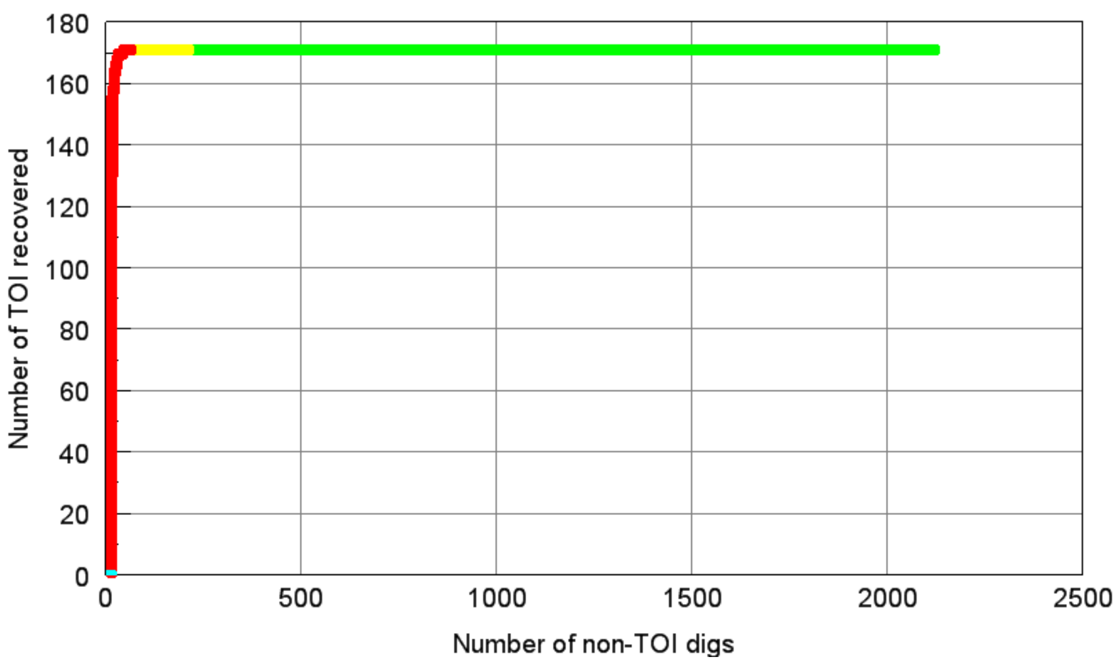


Figure 6-74. The ROC curve of the TEMTADS data using only the multi-dipole solver, new decision rules and fixing the polarity of coil 5.

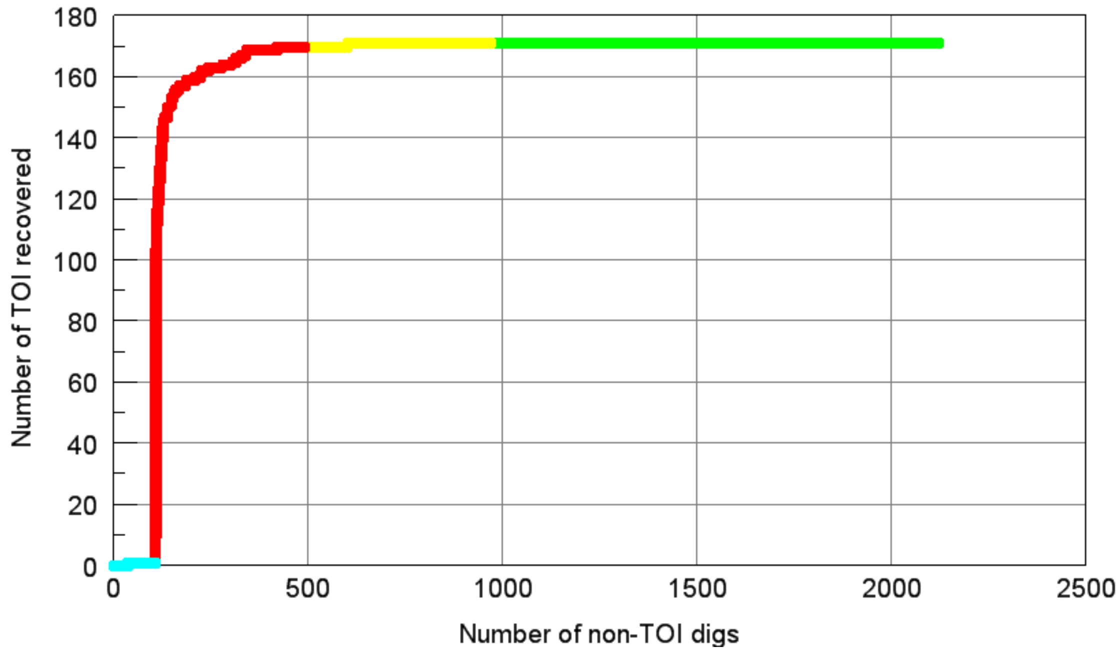


Figure 6-75. The ROC curve of the MM data using both the single and multi-dipole solvers, new decision rules and removing data from Rx2Y and Rx3Y.

7.0 COST ASSESSMENT

The data analysis performed in this demonstration consists of a number of distinct and sequential operations. The operations include anomaly extraction, characterization (parameter estimation), classification, and documentation. Table 7-1 reports forward-looking cost estimates for each of these operations, assuming a labor rate of $115 per hour and 2,000 anomalies. Costs for testing and comparing multiple inversion and classification schemes are not included in these costs.

All the dig lists using EM61 data were analyzed using UX-Analyze. We report estimated costs assuming only the EM61 data were analyzed and not costs using the EM61 data as a pre-screener to determine which anomalies required cued data. The cost of using the EM61 data as a pre-screener would not be much different than the EM61 only analysis costs. The same effort for anomaly extraction, characterization and classification would need to be performed. The only savings would come from documentation which is a small portion of the overall costs. The TEMTADS and Metal Mapper were analyzed using both UX-Analyze and IDL based analysis environment. The IDL based analysis ran only the single-dipole solver whereas the UX-Analyze methods ran both the single and multi-dipole solvers. Within each of the analysis environments, the TEMTADS and Metal Mapper used virtually the same process. The TEMTADS inversion routines ran slower because the TEMTADS system records more data due to the greater number of coils and time gates but that only affects computer processing time and not the analyst's time. The TEMTADS system also has a 2 square meter footprint compared to the 1 square meter for the Metal Mapper. The larger foot print and the monostatic terms allows the user the ability to visualize the data and select which coils will be used in the inversion to minimize the adverse affects secondary sources have on the data inversions. For these reasons the data extraction and parameter estimation costs are more for the TEMTADS system than the Metal Mapper. This is also the reason the IDL based data extraction costs are higher than the UX-Analyze based costs. Because UX-Analyze used the multi-dipole solver the analyst did not need to select coils to pass to the inversion.

The costs associated with "Data Extraction" and "Parameter Estimation" are reported per anomaly. It will not make much of a difference to these costs per anomaly if 100 anomalies are analyzed or 10,000 anomalies. The costs will just scale accordingly.

On the other hand the cost associated with "Classifier Training" and "Classification and Construction of the Ranked Anomaly List" are largely independent on the number of anomalies. Once the training data is analyzed and decision rules are developed the difference in time and cost to run 100 anomalies versus 10,000 anomalies through the classifier will be negligible because it is mainly based on computer speed and not operator dependent.

Table 7-1. Cost Summary by each individual data set.

Cost Category	**Description**	**Estimated Costs per Anomaly**
Processing Costs (Costs tracked by individual data set)		
Data Extraction	Cost/time required to extract data chip encompassing each anomaly	UXA EM61 Cart - $1 UXA TEMTADS - $0 UXA Metal Mapper - $0 IDL TEMTADS - $1.5 IDL Metal Mapper - $0
Parameter Estimation	Cost/time required to extract parameters for all anomalies and QC results.	UXA EM61 Cart - $2 UXA TEMTADS - $4 UXA Metal Mapper - $4 IDL TEMTADS - $7 IDL Metal Mapper - $6
Classifier Training	Cost/time required to optimize classifier design and train	UXA EM61 Cart - $1 UXA TEMTADS - $1.5 UXA Metal Mapper - $1.5 IDL TEMTADS - $0.5 IDL Metal Mapper - $0.5
Classification and Construction of the Ranked Anomaly List	Cost/time required to classify anomalies in the test set and construct the ranked anomaly list	UXA EM61 Cart - $.05 UXA TEMTADS - $.05 UXA Metal Mapper - $.05 IDL TEMTADS - $.05 IDL Metal Mapper - $0.05
Totals	Total cost to process and classify an anomaly.	UXA EM61 Cart - $4.05 UXA TEMTADS - $5.55 UXA Metal Mapper - $5.55 IDL TEMTADS - $9.05 IDL Metal Mapper - $6.55

7.1 COST DRIVERS

Data collection: Generally speaking, data collection costs will be greater for classification than detection only. Basically, the EMI analysis process utilizes subtle changes in the anomaly shape. Care must be taken during data collection to not only sample the anomaly fine enough, but also to not introduce noise due to inappropriate collection methods. The costs for data collection vary widely, depending on site conditions such as topography, vegetation, geologic background, known munitions types, as well as weather conditions.

Data Analysis: In general, data analysis costs will be greater for classification than detection only. Data analysis costs are affected by the presence of complex geology, which can make filtering and parameter estimation more complicated. The munitions of interest will also have a great effect on complexity and costs of processing, as will anomaly density. In the case considered here, there were three main TOI and a few areas with high anomaly density. Using the discrimination procedures described in this report and fixing problematic data we were able classify more than 60-90% of the non-munitions items correctly without any false negatives using the cued sensors. If we only used the dynamically collected EM61-MK2 data only 15% of the non-munitions items were correctly classified with a couple of false negatives. The number of non-munitions that can be removed with high confidence at another site may be much higher of lower depending on types of clutter and TOI. In addition, the job of the processor in determining the important features and training the classifier may be harder.

Excavation Cost. The costs associated with excavating anomalies vary widely and the goal is to reduce these costs via classification. Safety procedures and nominal burial depth drive remediation costs. When minimal engineering controls are used, costs as low as $45-90 per hole have been reported. When safety procedures are far more elaborate due either to the type of munitions or to their proximity to high value objects, the costs per hole are measured in the hundreds. With regards to burial depth, it is less costly to recover shallow, near-surface items than large deep targets.

7.2 COST BENEFIT

The cost benefit of the classification approach relates to savings realized by not excavating items that are not of interest. The ROC curve in Figure 7-1 shows a three-category classification scheme with a threshold set such that all the items on the right are high confidence non-TOI (Category 1). Although this is an example ROC only, it is very similar in nature to those presented in throughout this report. Note that the anomalies to the right of the threshold were correctly classified as high confidence not munitions. Cost savings can be realized, therefore, if we make use of the classification information and remediate accordingly.

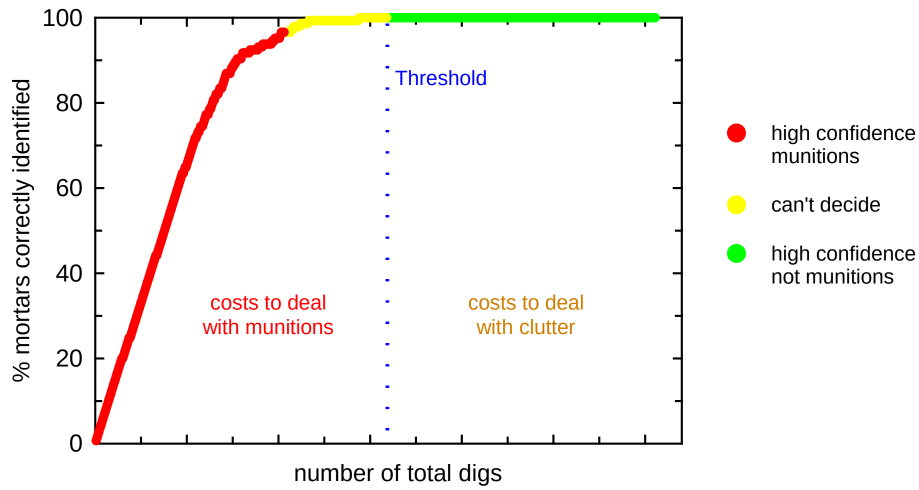


Figure 7-1. Example ROC curve to illustrate cost saving due to proper classification.

Figure 7-2 shows how notional costs accumulate through the process of data collection and processing, digging the munitions, and excavation. In the figure, the detection only (solid black line specifies a lower density data collection for detection only and all anomalies are excavated using intrusive recovery procedures that require trained UXO qualified personnel and safety equipment. The classification 1 (dashed green line) specifies higher density and quality data collection followed by classification processing, and all high-confidence clutter items are left unexcavated. Finally, the classification 2 (dotted green line) specifies higher density and quality data collection followed by classification processing, but a less expensive alternative to the current operational methods of intrusive recovery is used on the anomalies determined to be clutter with high confidence. The classification examples are tied to the different regions of the ROC curve shown in Figure 7-1. There are several important points to note in interpreting this curve: (i) The cumulative cost curves start out on the y-axis at different points. This reflects that the initial costs of higher density data collection and processing for classification are higher than the standard methods. The costs of digging the munitions, which must be borne in all cases, are included here. (ii) The "detection only" curve (solid black line) has a constant slope and ends at the total number of anomalies. All detected anomalies are dug using the same procedures at the same costs. (iii) For both classification examples, all of the items determined to be high confidence munitions or "can't decide" must be dug as though they are munitions. Thus, the two classification examples rise at a slope equal to the detection slope until the threshold is reached on the ROC curve where clutter is identified with high confidence (i.e., the yellow-green transition in Figure 7-1). (iv) In the region where there is high confidence that the remaining anomalies are clutter (green portion of the ROC curve) and it is decided not to dig these anomalies at all, no additional costs are incurred. (v) In the region where there is high confidence that the remaining anomalies are clutter and it is decided to dig these anomalies, but using alternative dig procedures, additional costs are incurred, but the cost of each of these digs is lower so the slope is more gradual. (vi) The break point in cost saving will be determined by the true dollars associated with the data collection, processing and excavation costs – all of which are site specific.

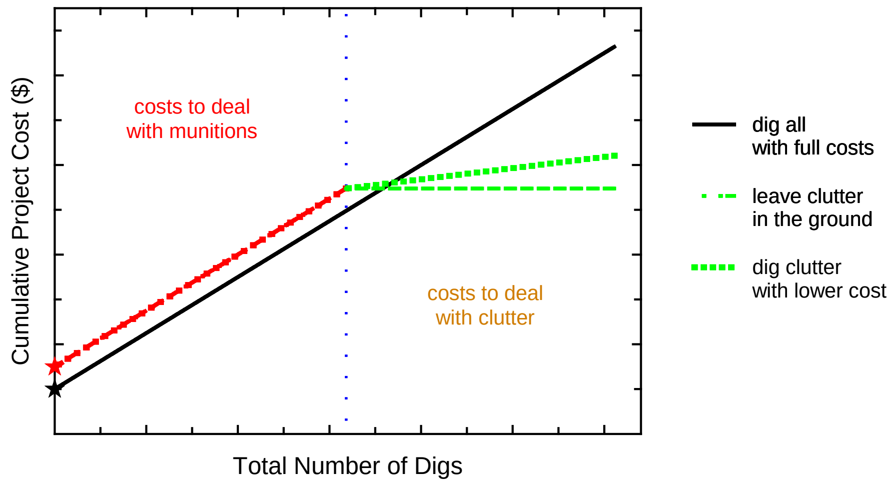


Figure 7-2. Conceptual cost model illustrating the potential savings using the classification methods outlined in this report.

Projecting forward based on experiences from this demonstration, the notional costs for classifying data similar in number and quality to that processed here are approximately $4 for EM61 data and $5.5 to $9 for the TEMTADS and Metal Mapper. There are additional costs associated with collecting higher density and quality data for the EM61 based system but the cost savings accrued by digging fewer anomalies is normally much greater. At this site however, the additional costs associated with collecting and analyzing the EM61 data were roughly on par with the savings associated with digging approximately 300 fewer anomalies. The TEMTADS and Metal Mapper are cued systems and only required the standard initial detection survey. The added data collections costs only arise from collecting the cued data which is estimated to be about $15 per anomaly. Adding this to the analysis costs results in additional costs of approximately $20.5-$24 per anomaly for the cued systems. When compared with excavation costs of $50-$100 per hole, the case for advanced classification is still cost effective because of greater number of anomalies that can be classified as non-TOI.

8.0 IMPLEMENTATION ISSUES

8.1 REGULATORY AND END-USER ISSUES

The ESTCP Program Office established an Advisory Group to facilitate interactions with the regulatory community and potential end-users of this technology. Members of the Advisory Group include representatives of the US EPA, State regulators, Corps of Engineers officials, and representatives from the services. The ESTCP staff worked with the Advisory Group to define goals for this Program and developed Project Quality Objectives. As the analyzed data from the demonstrations become available, the Advisory Group assisted in developing a validation plan.

The discrimination mindset that should be promoted is one that encourages geophysical service providers to deliver a dig list that is prioritized according objectives defined by the site's stakeholders. The decision metrics and thresholds should be quantitative, transparent, and documented. Stakeholder priorities could reasonably relate to size, depth of burial, shape, material type, or munitions type(s). Once the prioritized dig list is delivered, the stakeholders decide the ensuing actions. Stakeholders and/or site managers can realize financial savings by either modifying excavation procedures based upon the probability of being a TOI or by declaring that no further action is required for specific anomalies. The point is that the decision to take action always remains with the stakeholders – as it must, if the discrimination mindset is to be accepted.

9.0 REFERENCES

1. "MTADS Discrimination Array (TEMTADS) Demonstration Plan Supplement, Former Camp Butner," D.A. Steinhurst, submitted to ESTCP Program Office, May, 2010.

2. "Metal Mapper Data Collection Report: Camp Butner Discrimination Study" Mark Prouty, submitted to ESTCP Program Office, July, 2010.

3. "Shape-Based Classification and Discrimination of Subsurface Objects Using Electromagnetic Induction," Thomas H. Bell, Bruce Barrow and Nagi Khadr, Proceedings of the International Geoscience and Remote Sensing Symposium, Seattle, Washington, July 6-10, 1998.

4. "Model-Based Characterization of EM Induction Signatures for UXO/Clutter Discrimination Using the *MTADS* Platform" Bruce Barrow and H. H. Nelson, Proceedings of the UXO Forum 1999, Atlanta, Georgia, May 25-27, 1999.

5. "Source Separation using Sparse-Solution Linear Solvers" Jonathan T. Miller, Dean Keiswetter, Jim Kingdon, Tom Furuya, Bruce Barrow, and Tom Bell. *To be s*ubmitted to SPIE conference on Defense Security and Sensing, Orlando, FL, April, 2010.

Appendix A: Points of Contact

POINT OF CONTACT	ORGANIZATION	Phone Fax e-mail	Role in Project
Dr. Jeff Marqusee	ESTCP Program Office 901 North Stuart Street, Suite 303 Arlington, VA 22203	703-696-2120 (V) 703-696-2114 (F) jeffrey.marqusee@osd.mil	Director, ESTCP
Dr. Anne Andrews	ESTCP Program Office 901 North Stuart Street, Suite 303 Arlington, VA 22203	703-696-3826 (V) 703-696-2114 (F) anne.andrews@osd.mil	Deputy Director, ESTCP
Dr. Herb Nelson	ESTCP Program Office 901 North Stuart Street, Suite 303 Arlington, VA 22203	703-696-8726 (V) 703-696-2114 (F) 202-215-4844 (C) herbert.nelson@osd.mil	Program Manager, MM
Ms. Katherine Kaye	HydroGeoLogic, Inc. 11107 Sunset Hills Road, St 400 Reston, VA 20190	410-884-4447 (V) kkaye@hgl.com	Program Manager Assistant, MM
Dr. Shelley Cazares	Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311	703-845-6792 (V) 703-578-2877 (F) scazares@ida.ord	Performance Assessment
Dr. Dean Keiswetter	SAIC 120 Quade Drive Cary, NC 27513	919-677-1560 (V) keiswetterd@saic.com	SAIC Project Lead
Dr. Thomas Bell	SAIC 4001 North Fairfax Drive,4th Floor Arlington, VA 22203	703-312-6288 (V) bellth@saic.com	Quality Assurance Officer
Mr. Tom Furuya	SAIC 120 Quade Drive Cary, NC 27513	919-677-1563 (V) furuyat@saic.com	Data Analyst
Dr. Jim Kingdon	SAIC 4001 North Fairfax Drive,4th Floor Arlington, VA 22203	703-248-7706 (V) kingdonj@saic.com	Data Analyst
Dr. Nagi Khadr	SAIC 4001 North Fairfax Drive,4th Floor Arlington, VA 22203	(217) 819-0824 (V) Nagi.khadr@saic.com	Data Analyst

Appendix B: NAEVA's Analysis Report

Classification Memo for the MetalMapper Advanced Sensor

Three Criteria Considered for All Targets (3crit)

For this classification method it was determined that the confidence metric giving equal weight to size (b1), shape 1 (b2/b1) and shape 2 (b3/b1) would be the primary means of ranking. For targets falling within Category 2 the fit coherence was also considered in the ordering of items on the final prioritized dig list. The values used in this demonstration are summarized in Tables 1 and 2.

Table 1: Definitions for Fit Coherence (Fit_Coh[8] in Geosoft database):

System	Cannot Analyze	Low Coherence Fit	High Coherence Fit
Sky	0 → 0.70	0.70 → 0.90	0.90 → 1
Geometrics	0 → 0.55	0.55 → 0.85	0.85 → 1

Table 2: Definitions for Confidence Metric (Con_Metric_111[0] in Geosoft database):

System	High Confidence Clutter	Cannot Decide – Likely Clutter	Cannot Decide – Likely TOI	High Confidence TOI
Sky	0 → 0.65	0.65 → 0.75	0.75 → 0.85	0.85 → 1
Geometrics	0 → 0.60	0.60 → 0.70	0.70 → 0.80	0.80 → 1

The high confidence clutter boundary was derived from an examination of the confidence metrics calculated for clutter items within the training data set. There were sufficient examples of clutter items here to determine a reasonable clutter threshold. The training set contained a relatively small number of targets of interest and the test pit provided several additional measurements. The target parameters for these items were added to the classification library and thus could not be used to determine the high confidence TOI threshold. Instead a value slightly below where the confidence metrics appear to stabilize near 1 within the test set was selected as the threshold.

The four basic classification categories were used as the first step of the sorting process. Classes were added to further refine target ranking.

- If the confidence metric was less than 0.00001 it was set to zero

Targets were placed into Category 4 if they met one or more of the following criteria:

- The fit location was over 1m from the flag location
- The fit location was over 0.65m from the center of the array
- The fit coherence fell below the Cannot Analyze boundary set in Table 1
- The target could not be fit to any item in the training library, the confidence metric was equal to zero

- The target was not successfully fit to a model during data inversion

Within Category 4 two classes were used to identify which part of the modeling and classification process caused the analysis failure.

- Targets that did not fit to a model during data inversion, and as a result could not be considered for classification, were assigned to Class 6.
- All other targets labeled as Cannot Analyze were assigned to class 5.

Targets were placed into Category 3 if they met the high confidence TOI criteria outlined in Table 2.

Targets were placed into Category 1 if they met the high confidence clutter criteria outlined in Table 2.

Targets were placed into Category 2 if they met the cannot decide criteria outlined in Table 2.

Targets in Category 2 were then assigned classes based on their fit coherencies and confidence metrics as defined in Tables 1 and 2.

- Class 1 – High Coherence Fit, Cannot Decide – Likely Clutter
- Class 2 – Low Coherence Fit, Cannot Decide – Likely Clutter
- Class 3 – Low Coherence Fit, Cannot Decide – Likely TOI
- Class 4 – High Coherence Fit, Cannot Decide – Likely TOI

Table 3: Description of classes and categories used for prioritization

Description	**Class**	**Category**
Clutter (High Confidence)		1
Cannot Decide (Possible Clutter)	1	2
Cannot Decide (Poor Fit - Possible Clutter)	2	2
Cannot Decide (Poor Fit - Possible TOI)	3	2
Cannot Decide (Possible TOI)	4	2
TOI (High Confidence)		3
Cannot Analyze	5	4
Cannot Analyze (Not Fit During Inversion)	6	4

Sorting was performed using three channels, with Category as the primary basis, Class as the secondary basis and the Decision Statistic (1-Confidence Metric) as the tertiary basis. A rank channel was then created and categories 1 through 3 were assigned the rank of 1+their fiducial number. The decision statistic and rank for all targets falling in category 4 was set to -9999.

The target list was then exported from Oasis and duplicate targets were removed using sorting and the application of math functions in Microsoft Excel. When duplicate targets were found (either multiple collections by one system or collections by both systems) the target with the highest rank was retained. This was an attempt to conservatively classify targets and reduce the number of TOI incorrectly classified as clutter. All targets falling into Category 4 were removed if there was an additional collection that fell into one of the other categories.

Table 4: Summary of targets within each category

<table>
<tr><th>Category</th><th># Anomalies</th><th>Description</th><th></th></tr>
<tr><td>1</td><td>1344</td><td>Can Analyze: Likely Non-TOI (Clutter)</td><td>Do Not Dig</td></tr>
<tr><td>2</td><td>229</td><td>Can Analyze: Cannot Decide</td><td rowspan="3">Dig</td></tr>
<tr><td>3</td><td>176</td><td>Can Analyze: Likely TOI</td></tr>
<tr><td>4</td><td>363</td><td>Cannot Analyze</td></tr>
</table>

Decision Memo for the MetalMapper Advanced Sensor

Two Stage Classification using Three or Two Criteria (3crit2crit)

For this classification method built in functions from the UX-Analyze module for Geosoft's OasisMontaj were used to model the data, calculate target parameters and match the parameters to a classification library. A two staged approach using two separate metric weighting methods to fit the target signatures to a library was used to rank the targets. The first stage used a 3-critera confidence metric generated by giving equal weight to the size (b1), shape 1 (b2/b1) and shape 2 (b3/b1) metrics. For anomalies with lower confidence metrics a second calculation was performed to generate a 2-critera confidence metric with a weight of 2 for the size (b1) and 1 for shape 1 (b2/b1). A combination of the fit coherence which measures the closeness of the modeled data to the measured data and the confidence metrics was used to generate the ranked dig list. The confidence metrics represent a measure of the quality of the fit of target parameters between the selected anomaly and a library containing example measurements of targets of interest (TOI) expected at the site. Due to the differences in the noise level between the two systems that were deployed at the site different parameter ranges were assigned to the two systems. The values used for parameters are summarized in Tables 1-3.

Table 1: Definitions for Fit Coherence (Fit_Coh[8] in Geosoft database)

System	Cannot Analyze	Low Coherence Fit	Mid Coherence Fit	High Coherence Fit
Sky	0 → 0.70	0.70 → 0.80	0.80 → 0.90	0.90 → 1
Geom	0 → 0.55	0.55 → 0.70	0.70 → 0.85	0.85 → 1

Table 2: Definitions for 3-Critera Confidence Metric (Con_Metric_111[0] in Geosoft database)

System	High Confidence Clutter	Cannot Decide – Possible Clutter	Cannot Decide – Possible Clutter	High Confidence TOI
Sky	0 → 0.65	0.65 → 0.75	0.75 → 0.85	0.85 → 1
Geom	0 → 0.60	0.60 → 0.70	0.60 → 0.70	0.80 → 1

Table 3: Definitions for 2-Critera Confidence Metric (Con_Metric_210[0] in Geosoft database)

System	High Confidence Clutter	Cannot Decide – Possible Clutter	Cannot Decide – Possible Clutter	High Confidence TOI
Sky	0 → 0.70	0.70 → 0.775	0.775 → 0.85	0.85 → 1
Geom	0 → 0.60	0.60 → 0.725	0.725 → 0.80	0.85 → 1

The high confidence clutter boundary was derived from an examination of the confidence metrics for the clutter items within the standard training set. There were sufficient examples of clutter items here to determine a reasonable clutter threshold. The training set contained a relatively small number of targets of interest and the test pit provided several additional measurements. The target parameters for these items were added to the classification library and thus could not be used to determine the high confidence TOI threshold. Instead a value slightly below where the confidence metrics appear to stabilize near 1 within the test set was selected as the threshold.

Ranking was performed using a combination of classes which were bound by the parameters listed above and within each class the anomalies were sorted by the confidence metric. The procedure for ranking the targets used a series of math expressions to identify the confidence metric to use, the class to place the anomaly in and to identify any targets that cannot be analyzed. An outline of the procedure follows:

- If the confidence metric is less than 0.00001 set it to zero

Cannot Analyze Category

- Targets where the modeling fit process failed are placed into class 10
- Targets where the flag location is greater than 1m from the fit location are placed in class 9
- Targets where the instrument location is greater than 0.65m from the fit location are placed in class 9
- Targets where the fit coherence falls below the Low Confidence threshold are placed in class 9
- Targets where both the 2 and 3 criteria confidence metrics are zero are placed in class 9

Determine Which Metric to Use

- If a target does not match to the library with 3-criteria use the 1-criteria fit
- If a target does not match to the library with 2-critiera use the 3-criteria fit
- If b3 does not decay as expected use the 2-criteria fit (which does not use the b3 parameter)
- If the 3-criteria fit does not place the target into the high confidence clutter or TOI category use the higher of the two confidence metrics

Can Analyze: Likely Non-TOI (Clutter)

- Using 3-criteria fits, targets with high fit coherence and low confidence metrics are placed into class 1
- Using 2-criteria fits, targets with high fit coherence and low confidence metrics are placed into class 2

Can Analyze: Likely TOI

- Using 3-criteria fits targets with high fit coherence and low confidence metrics placed into class 8
- Using 2-criteria fits targets with high fit coherence and low confidence metrics placed into class 7

Can Analyze: Cannot Decide

- Targets with a lower fit coherence are placed in category 2 and split into four classes
- Mid coherence fit and low confidence metric placed into class 3
- Low coherence fit and low confidence metric placed into class 4
- Low coherence fit and high confidence metric placed into class 5
- Mid coherence fit and high confidence metric placed into class 6

The decision statistic was set to the 1 minus the corresponding confidence metric based on if the 2 or 3-criteria fit was used. Ranking was performed by sorting using three channels with the Category (ascending) as the primary basis, Class (ascending) as the secondary basis and the Decision Statistic (ascending) as the tertiary basis. The rank value was assigned as increasing integers to the sorted list. The decision statistic and rank for all targets falling in category 4 was set to -9999.

Table 4: Classification with equal weight 3-critera and unequal weight 2-critera

Description	**Class**	**Category**
Clutter (High Conf 3-critera)	1	1
Clutter (High Conf 2-critera)	2	1
Cannot Decide (possible clutter)	3	2
Cannot Decide (poor fit possible clutter)	4	2
Cannot Decide (poor fit possible TOI)	5	2
Cannot Decide (possible TOI)	6	2
TOI (High Conf 2-critera)	7	3
TOI (High Conf 3-critera)	8	3
Cannot Analyze	9	4
Cannot Analyze (did not fit)	10	4

The target list was then exported from Oasis and duplicate targets were removed in Excel. When duplicate targets were identified (either multiple collections by one system or collected by both systems) the target with the highest rank was retained. This was an attempt to conservatively classify targets and reduce the number of possible TOI incorrectly classified as clutter. All duplicate targets falling into category 4 were removed if there was an additional collection that fell into one of the other categories.

Table 5: Summary of targets within each category

<table>
<tr><td>Category</td><td># Anomalies</td><td>Description</td><td></td></tr>
<tr><td>1</td><td>1115</td><td>Can Analyze: Likely Non-TOI (Clutter)</td><td>Do Not Dig</td></tr>
<tr><td>2</td><td>478</td><td>Can Analyze: Cannot Decide</td><td rowspan="3">Dig</td></tr>
<tr><td>3</td><td>173</td><td>Can Analyze: Likely TOI</td></tr>
<tr><td>4</td><td>346</td><td>Cannot Analyze</td></tr>
</table>

Decision Memo for the MetalMapper Advanced Sensor

Two Stage Classification using Three or One Criteria (3crit1crit)

For this classification method built in functions from the UX-Analyze module for Geosoft's OasisMontaj were used to model the data, calculate target parameters and match the parameters to a classification library. A two staged approach using two separate metric weighting methods to fit the target signatures to a library was used to rank the targets. The first stage used a 3-critera confidence metric generated by giving equal weight to the size (b1), shape 1 (b2/b1) and shape 2 (b3/b1) metrics. For anomalies with lower confidence metrics a second calculation was performed to generate a 1-critera confidence metric using the size (b1). A combination of the fit coherence, which measures the closeness of the modeled data to the measured data and the confidence metrics was used to generate the ranked dig list. The confidence metrics represent a measure of the quality of the fit of target parameters between the selected anomaly and a library containing example measurements of targets of interest (TOI) expected at the site. Due to the differences in the noise level between the two systems that were deployed at the site different parameter ranges were assigned to the two systems. The values used for parameters are summarized in Tables 1-3.

Table 1: Definitions for Fit Coherence (Fit_Coh[8] in Geosoft database)

System	Cannot Analyze	Low Coherence Fit	Mid Coherence Fit	High Coherence Fit
Sky	0 → 0.70	0.70 → 0.80	0.80 → 0.90	0.90 → 1
Geometrics	0 → 0.55	0.55 → 0.70	0.70 → 0.85	0.85 → 1

Table 2: Definitions for 3-Critera Confidence Metric (Con_Metric_111[0] in Geosoft database)

System	High Confidence Clutter	Cannot Decide – Possible Clutter	Cannot Decide – Possible Clutter	High Confidence TOI
Sky	0 → 0.65	0.65 → 0.75	0.75 → 0.85	0.85 → 1
Geometrics	0 → 0.60	0.60 → 0.70	0.60 → 0.70	0.80 → 1

Table 3: Definitions for 1-Critera Confidence Metric (Con_Metric_100[0] in Geosoft database)

System	High Confidence Clutter	Cannot Decide – Possible Clutter	Cannot Decide – Possible Clutter	High Confidence TOI
Sky	0 → 0.75	0.75 → 0.85	0.85 → 0.95	0.95 → 1
Geometrics	0 → 0.70	0.70 → 0.8125	0.8125 → 0.925	0.925 → 1

The high confidence clutter boundary was derived from an examination of the confidence metrics for the clutter items within the standard training set. There were sufficient examples of clutter items here to determine a reasonable clutter threshold. The training set contained a relatively small number of targets of interest and the test pit provided several additional measurements. The target parameters for these items were added to the classification library and thus could not be used to determine the high confidence TOI threshold. Instead a value slightly below where the confidence metrics appear to stabilize near 1 within the test set was selected as the threshold.

Ranking was performed using a combination of classes which were bound by the parameters listed above and within each class the anomalies were sorted by the confidence metric. The procedure for ranking the targets used a series of math expressions to identify the confidence metric to use, the class to place the anomaly in and to identify any targets that cannot be analyzed. An outline of the procedure follows:

- If the confidence metric is less than 0.00001 set it to zero

Cannot Analyze Category

- Targets where the modeling fit process failed are placed into class 10
- Targets where the flag location is greater than 1m from the fit location are placed in class 9
- Targets where the instrument location is greater than 0.65m from the fit location are placed in class 9
- Targets where the fit coherence falls below the Low Confidence threshold are placed in class 9
- Targets where both the 1 and 3 criteria confidence metrics (this indicates that the target could not be fit to any item in the training library) are zero are placed in class 9

Determine Which Metric to Use

- If a target does not match to the library with 3-criteria use the 1-criteria fit
- If a target does not match to the library with 1-critiera use the 3-criteria fit
- If b2 or b3 does not decay as expected use the 1-criteria fit (which does not use these parameters)
- If the 3-criteria fit does not place the target into the high confidence clutter or TOI category use the higher of the two confidence metrics

Can Analyze: Likely Non-TOI (Clutter)

- Using 3-criteria fits, targets with high fit coherence and low confidence metrics are placed into class 1
- Using 1-criteria fits, targets with high fit coherence and low confidence metrics are placed into class 2

Can Analyze: Likely TOI

- Using 3-criteria fits targets with high fit coherence and low confidence metrics placed into class 8
- Using 1-criteria fits targets with high fit coherence and low confidence metrics placed into class 7

Can Analyze: Cannot Decide

- Targets with a lower fit coherence are placed in category 2 and split into four classes
- Mid coherence fit and low confidence metric placed into class 3
- Low coherence fit and low confidence metric placed into class 4
- Low coherence fit and high confidence metric placed into class 5
- Mid coherence fit and high confidence metric placed into class 6

The decision statistic was set to one minus the corresponding confidence metric based on if the 1 or 3-criteria fit was used. Ranking was performed by sorting using three channels with the Category (ascending) as the primary basis, Class (ascending) as the secondary basis and the Decision Statistic (ascending) as the tertiary basis. The rank value was assigned as increasing integers to the sorted list. The decision statistic and rank for all targets falling in category 4 was set to -9999.

Table 4: Classification with equal weight 3-critera and unequal weight 2-critera

Description	Class	Category
Clutter (High Conf 3-critera)	1	1
Clutter (High Conf 1-critera)	2	1
Cannot Decide (possible clutter)	3	2
Cannot Decide (poor fit possible clutter)	4	2
Cannot Decide (poor fit possible TOI)	5	2
Cannot Decide (possible TOI)	6	2
TOI (High Conf 1-critera)	7	3
TOI (High Conf 3-critera)	8	3
Cannot Analyze	9	4
Cannot Analyze (did not fit)	10	4

The target list was then exported from Oasis and duplicate targets were removed in Excel. When duplicate targets were identified (either multiple collections by one system or collected by both systems) the target with the highest rank was retained. This was an attempt to conservatively classify targets and reduce the number of possible TOI incorrectly classified as clutter. All duplicate targets falling into category 4 were removed if there was an additional collection that fell into one of the other categories.

Table 5: Summary of targets within each category

Category	# Anomalies	Description	
1	1104	Can Analyze: Likely Non-TOI (Clutter)	Do Not Dig
2	435	Can Analyze: Cannot Decide	Dig
3	230	Can Analyze: Likely TOI	
4	343	Cannot Analyze	

Decision Memo for the MetalMapper Advanced Sensor

Two Stage Classification using Three or One Criteria with Multi Source Solver (3crit1critMulti)

For this classification method the completed 3crit1crit prioritization using the single target solver was used as the starting point. Targets that had been placed in the Cannot Decide and Cannot Analyze categories were then run through the UX-Analyze multi source solver to try and reduce the number of targets falling into these categories. A total of 778 targets were modeled with the multi source solver and the remaining 1334 targets the parameters from the single source solver will continue to be used for classification. This list was not submitted for scoring, a self analysis of the results was done base on previously provided ground truth.

The same classification method for the 3crit1crit dig list was used. The target list was then exported from Oasis and duplicate targets were removed in Excel. When duplicate targets were identified (either multiple collections by one system or collected by both systems) the target with the highest rank was retained. This was an attempt to conservatively classify targets and reduce the number of possible TOI incorrectly classified as clutter. All duplicate targets falling into category 4 were removed if there was an additional collection that fell into one of the other categories. The use of the multi source solver does improve the results however there are still a relatively high number of Cannot Analyze targets – this is based mainly on the fit coherence that was used. Summary target counts for the single vs multi source solvers are shown in the two tables below. Overall the number of targets to dig (in is slightly reduced from 1008 to 870, reduction of 138 about 14%.

Table 1: Summary of targets within each category

<table>
<tr><td>Category</td><td>Single Solver

Anomalies</td><td>Mulit Solver # Anomalies</td><td>Description</td><td></td></tr>
<tr><td>1</td><td>1104</td><td>1242</td><td>Can Analyze: Likely Non-TOI (Clutter)</td><td>Do Not Dig</td></tr>
<tr><td>2</td><td>435</td><td>440</td><td>Can Analyze: Cannot Decide</td><td rowspan="3">Dig</td></tr>
<tr><td>3</td><td>230</td><td>238</td><td>Can Analyze: Likely TOI</td></tr>
<tr><td>4</td><td>343</td><td>192</td><td>Cannot Analyze</td></tr>
<tr><td>5</td><td>179</td><td>179</td><td>Training Data</td><td>Dig</td></tr>
<tr><td>-</td><td>1187</td><td>1049</td><td>Targets below dig line (adds in 179 training digs)</td><td></td></tr>
</table>

Table 2: Summary of targets within each category

Single Solver # Anomalies	Mulit Solver # Anomalies	Description

0	0	TOI in Category 1
3	6	TOI in Category 2
131	132	TOI in Category 3
5	1	TOI in Category 4
6	6	TOI in Category 5
1157	951	# Digs to 100% TOI (including 179 training digs)

Appendix C: Parsons' Analysis Report

DISCRIMINATION TEST REPORT
FORMER CAMP BUTNER

Durham, Granville and Person Counties, North Carolina

prepared for:

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

Subcontract P010056938

prepared by:

PARSONS

1700 Broadway
Denver, Colorado 80290

Revision No. 0

March 2011

Craig Murray

Project Geophysicist: ______ FOR CRAIG MURRAY ______ 3/30/11

(Signature) **(Date)**

Greg Van

Project Manager: ______ Greg Van ______ 3/30/11

(Signature) **(Date)**

This page intentionally left blank

INTRODUCTION

Parsons performed a number of data processing and analysis tasks as part of the UXO Classification Study at the former Camp Butner in Durham, Granville, and Person Counties, N.C. This report describes the procedures used to process data collected at the site and supplied to Parsons by Science Applications International Corporation (SAIC), Parsons' experience using the UX-Analyze software extension of Geosoft's Oasis Montaj, the technical approach used to create ranked dig lists for the project, and includes an analysis of the reasons the lists failed to accurately discriminate targets of interest (TOI) from non-TOI.

UX-ANALYZE DATA PROCESSING

EM61-MK2 Target Inversion

Parsons received a master database containing the EM61-MK2 data for the Camp Butner project site and an Excel spreadsheet containing the full set of targets selected from the data. The first step used in processing the data was inverting the selected targets to determine modeled parameters for the location, size, and orientation of the source object, as well as the polarizability of each axis of the object. The difference between the expected model and actual model calculated for each object was also calculated during this step.

All initial inversion was performed using the UX-Analyze batch processing mode and window size of 5 meters (m). Following initial batch processing, individual targets were examined using the interactive mode. Polygons were drawn by hand for each individual target as necessary to mask out adjacent anomalous data within the 5m window. Hand drawing polygons for the targets was particularly time consuming, although subsequent experience with UX-Analyze suggests that using the UXAPolyCreate.GX would have saved a considerable amount of time in this step.

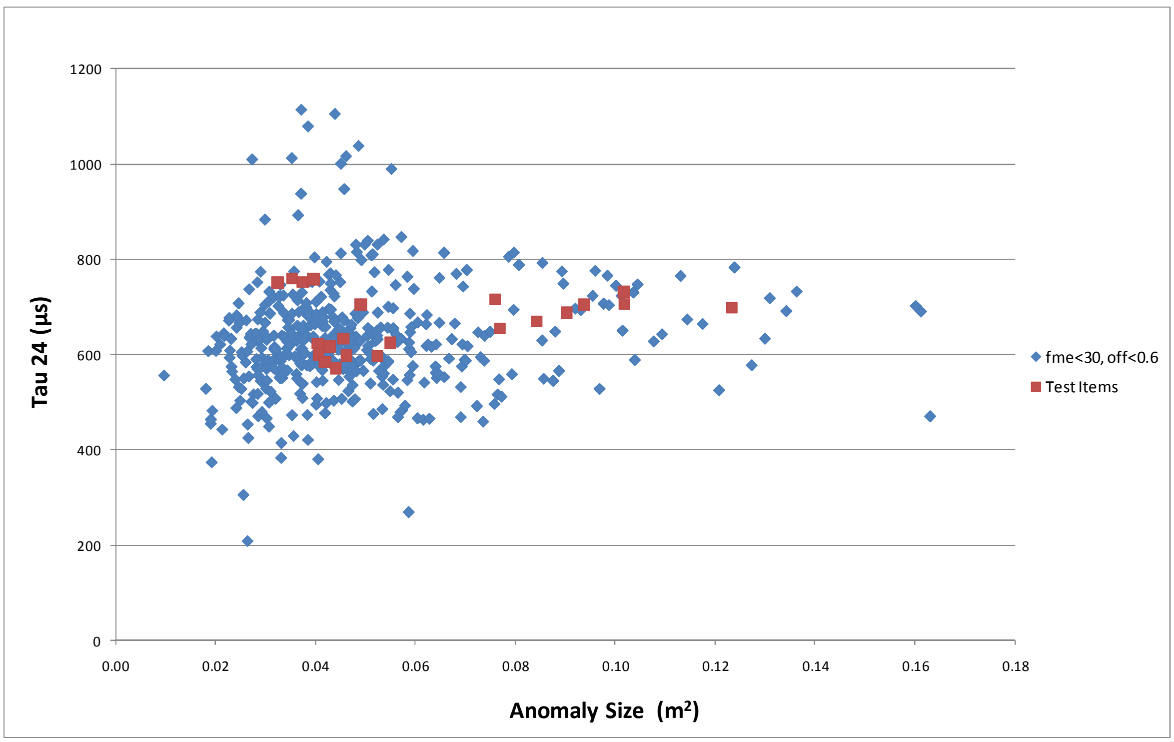
The final step in determining parameters for the individual EM61-MK2 targets was determining decay values. Decay values were calculated for all time gates combinations (i.e., 1-2, 1-3, 1-4, 2-3, 2-4, and 3-4) using both the UXADecay.GX function and a .GX developed by Parsons. The two .GXs use different methods to determine an anomaly's peak and decay value, and both were used to create ranked dig lists for a comparison study of discrimination results.

Initial processing of EM61-MK2 targets also included determining a list of Cannot Analyze targets. These included targets in clusters, or those for which the response of adjacent targets seemed to be affecting the response of the target in question. It was assumed that overlapping responses would negatively affect both the calculation of decay constants and the use of Metal Mapper results to discriminate between TOI and non-TOI. The Cannot Analyze list included 263 targets.

Training Data Selection

The modeled results for each EM61-MK2 target were used to determine initial assumptions regarding the likelihood that specific targets would be related to TOI. These assumptions used a combination of the fit model error and size results generated during the inversion of the EM61-MK2 targets to determine a subset of targets considered the most likely to be TOI. This subset of targets, which included those with fit model errors less than 30 percent and offsets less than 0.6 meters (m), were graphed with the Camp Butner test plot data to determine if additional discrimination could be performed using either the anomaly size or decay values calculated during inversion of the EM61-MK2 targets. The resulting graph (**Figure 1**) indicated that thresholds potentially could be drawn for discrimination purposes at size greater than $0.6m^2$ and Tau 2-4 values less than 500 microseconds (µs) and greater than 800µs.

Figure 1
Decay Constant (Tau 2-4) vs. Target Size
for Targets Considered Most Likely to be TOI

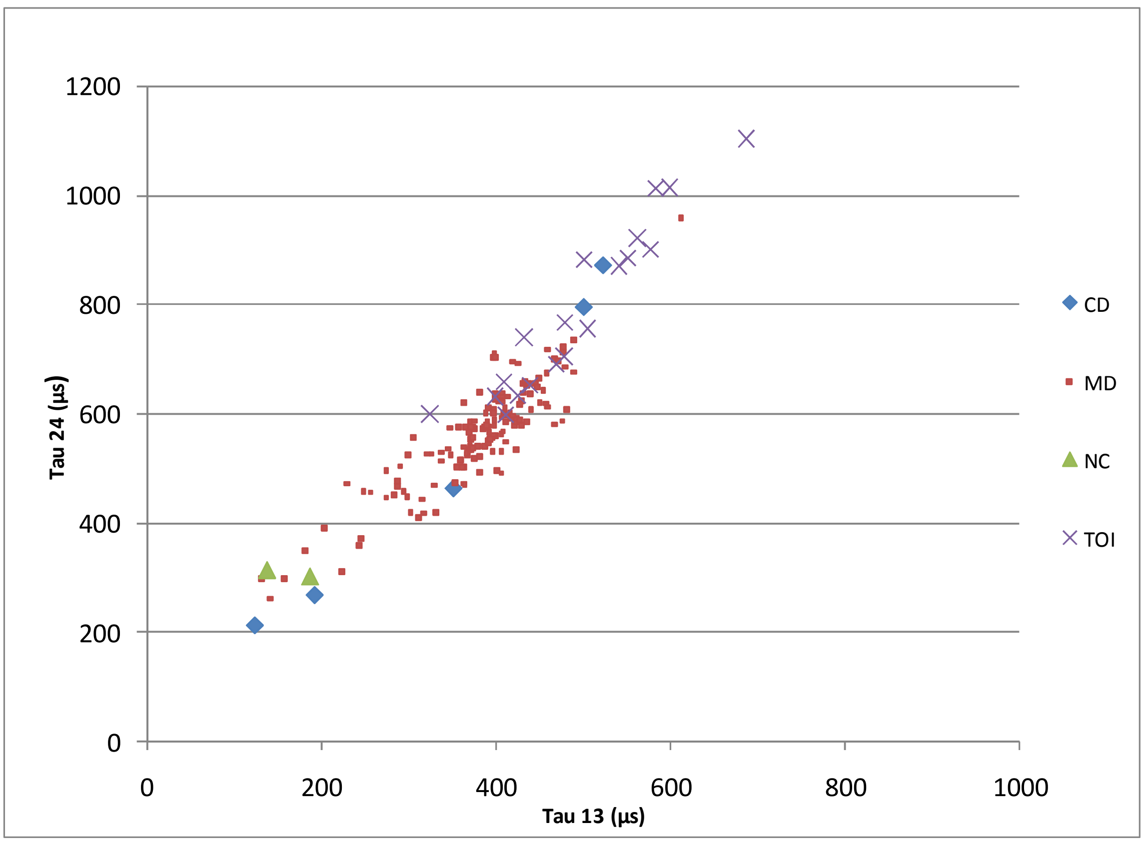


A custom training data set of 181 targets was generated using these assumptions. In general, a small number of training targets were selected to confirm or disprove the assumptions that large targets (i.e., greater than $0.6m^2$) were more likely to be caused by TOI and that targets outside of the Tau 2-4 decay constant range of 500µs to 800µs were unlikely to be caused by TOI. A much larger subset of training targets was selected from the remaining targets to determine if parameters could be generated to discriminate within the large grouping of targets with sizes smaller than $0.6m^2$ and within the Tau 2-4 decay constant range of 500 µs to 800µs.

Training Data Results

The training data results were separated according to dig type (no contact [NC], cultural debris [CD], munitions debris [MD], and TOI) and graphed using various parameters to attempt to determine discrimination rules using size and decay constant prior to requesting a cued data set for use in additional discrimination analysis. This analysis suggested that the largest differentiator between TOI and non-TOI for the Camp Butner targets would be decay values, specifically Tau 1-3 and Tau 2-4 (**Figure 2**). Based on the limited amount of training data available, Tau 2-4 seemed best suited for discriminating TOI from non-TOI in the data set. Prior to requesting cued Metal Mapper data, an initial discrimination rule was developed to consider targets with Tau 2-4 values greater than 800µs indicative of TOI and targets with Tau 2-4 values less than 500µs indicative of non-TOI.

Figure 2
Decay Constant (Tau 2-4) vs. Decay Constant (Tau 1-3)
for Training Data



Using this initial discrimination criteria, 443 targets were considered to be non-TOI (Tau 1-3 < 500µs), and 50 targets were considered likely to be TOI (Tau 2-4 ≥ 800µs). Cued Metal Mapper data were requested for the 1,578 targets not already considered TOI, non-TOI, or Cannot Analyze.

Metal Mapper Target Inversion

Parsons requested cued Metal Mapper data for the 1,578 targets not already classified as non-TOI or likely TOI based on their decay values or as Cannot Analyze. The supplied cued data included results for 1,834 targets, which included duplicate data for targets where re-collections were performed. Cued data were collected using two different Metal Mappers, one by Sky Research and one by Geometrics. The data collected by the two Metal Mappers were processed separately based on configuration differences between the units, although the same processing steps were used for both data sets.

All target inversion was performed using the UX-Analyze batch processing mode with the multiple object solver enabled. Initially, an attempt was made to process all of the data collected using each Metal Mapper (Sky and Geometrics) in one database; however, a number of targets at the end of the list were not inverted using this method. Although the exact cause is unknown, it is possible that the processing computer went into standby mode during the inversion and that processing stopped afterwards. Target lists were subsequently broken down into groups of approximately 150 for inversion, and no additional problems were noted.

The UX-Analyze advanced target analysis feature was used to determine the most likely source objects for the cued data targets. Analyses were performed for each target using four different library databases:

- the full library supplied by SAIC,
- the full library with 20 millimeter (mm) projectiles removed,
- a library containing only clutter and the TOI possibly present at Camp Butner (i.e., 37mm, 40mm, 81mm, 105mm, 155mm, M48 fuzes, and 2.36-inch rockets), and
- a library containing only the Camp Butner munitions (clutter removed).

The metric weights for each polarization curve (size, shape 1, and shape 2) were left at the default 1:1:1 for all analysis.

TARGET DISCRIMINATION

This section discusses the two approaches Parsons used for target discrimination: classification using combinations of EM61-MK2 decay values and inverted Metal Mapper data, and classification using EM61-MK2 decay values alone. Each classification approach will be discussed separately, including the selection of parameters, categorization of targets, creation of ranked dig lists, and results.

Classification by EM61-MK2 and Metal Mapper Data

Parameters

Prior to the development of separate approaches, Parsons reviewed the 181-item set of training data and the 22-item set of test plot data to identify parameters that could be used

to classify the master list of 2,113 targets (exclusive of training targets). As discussed in the Training Data Selection section, initial classification rules were developed using decay constants alone as determined during EM61-MK2 target inversion; and the cued data request did not include targets already classified using decay values. Following inversion of the Metal Mapper targets, additional rules were determined using the results of the UX-Analyze advanced target classification analysis, specifically the confidence matches and confidence metrics generated by this analysis.

For the Metal Mapper dig lists compiled, metrics were developed to classify targets based on a combination of decay values and the confidence metric and confidence match results from the advanced target classification analysis. Decay constant results for channels 2 to 4 were used to classify two groups of targets as either non-TOI (Tau 2-4 < 500µs) or TOI (Tau 2-4 > 800µs) based solely on the results of the EM61-MK2 analysis. The ten highest confidence metric/confidence match results from the advanced target analysis performed for each Metal Mapper target were then examined for the remaining targets. Two subsets of results were identified in this analysis: 1) targets for which at least one of the ten results indicated that the target was a potential TOI and 2) targets for which all ten results indicated that the target was some sort of clutter. Targets were placed into one of the following four categories for discrimination purposes:

Can Analyze – Likely Clutter

These targets were classified based on a very low decay constant, a high probability of being clutter, or a low probability or being a TOI. A target is assigned to Category 1: Can Analyze – Likely Clutter if its Tau 2-4 value was less than 500µs, if all of the 10 confidence matches indicated that it was some sort of clutter, or if the confidence metrics for all matches indicating possible TOI were below 0.60. Its rank within Category 1 is based on the following subsets, from lowest to highest:

1) Confidence matches from advanced target analysis that included only Clutter results for all 10 matches, with higher confidence matches receiving lower ranks;

2) Tau 2-4 values below 500µs, with lower decay values receiving lower ranks

3) At least one confidence match from advanced target analysis indicates possible TOI, but all TOI confidences are below 0.600, with lower TOI confidences receiving lower ranks.

Can Analyze – Cannot Decide

These targets were classified based on at least one confidence match indicating a possible TOI, but not with enough confidence that it was very likely to be a TOI. The only targets assigned to Category 2: Cannot Decide were those for which the advanced target classification indicated a possible TOI with a confidence of at least 0.600 but less than 0.750. Lower TOI confidences received lower ranks.

Can Analyze – Likely Target of Interest

These targets were classified based on very high decay constant or a high probability of being a TOI. A target is assigned to Category 3: Likely TOI if its Tau 2-4 value was 800µs or higher or if the confidence metric for one of the matches indicating possible TOI was above 0.750. Its rank within Category 3 is based on the following subsets, from lowest to highest:

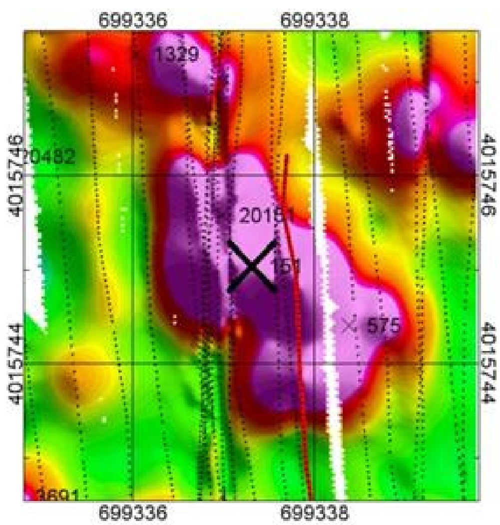
1) Tau 2-4 values greater than 800µs, with lower decay values receiving lower ranks

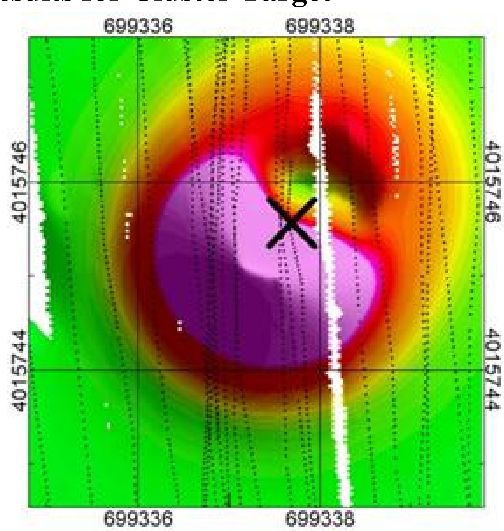
2) At least one confidence match from advanced target analysis indicates possible TOI with a confidence of 0.750 or higher, with lower TOI confidences receiving lower ranks.

Cannot Analyze

Cannot Analyze targets were identified prior to any inversion as those for which the response of adjacent EM61-MK2 targets seemed to be affecting the response of the target in question. It was assumed that overlapping responses would negatively affect the calculation of EM61-MK2 decay constants and the use of Metal Mapper results to discriminate between TOI and non-TOI. The Cannot Analyze list included 263 targets and was the same for all ranked lists submitted during the project. Figure 3 shows an example of poor modeling due to the effects of other, nearby targets.

Figure 3
Actual and Modeled Results for Cluster Target





Decision Statistic

A Decision Statistic was calculated for each target after it had been assigned a Category. Decision Statistic values range from 0.000 (most likely clutter) to 4.000 (most likely munition). For the four lists generated using cued Metal Mapper data, the following rules were used to compute the decision statistic:

1) Category 1: Can Analyze – Likely Clutter (0 – 2.599)

   1 – confidence metric for targets with only clutter results from advanced target classification (0 – 1.000)

   1 + tau24 * 0.001 for targets with tau24 results < 500μs (1.000 – 1.500)

   2 + confidence metric for targets with at least one TOI result from advanced target classification (2.000 – 2.599)

2) Category 2: Can Analyze – Cannot Decide (2.600 – 2.749)

   2 + confidence metric for targets with at least one TOI result from advanced target classification

3) Category 3: Can Analyze – Likely TOI (2.80 – 4.000):

   2 + tau24 * 0.001 for targets with tau24 results > 800μs (2.800 – 3.750)

   3 + confidence metric for targets with at least one TOI result from advanced target classification (3.750 – 4.000)

4) Cannot Analyze:

   Those targets classified as "Cannot Analyze" were assigned a Decision Statistic of -9999, according to the directions provided to demonstrators in this study.

Ranked Dig Lists

Four ranked dig lists were compiled using the Categories and Decision Statistics described above. For each list, targets were categorized and ranked according to the Decision Statistic computed for each target. Where applicable, targets with identical Decision Statistics were prioritized according to the channel 2 responses, with higher responses considered more likely to be TOI. Table 1 contains a brief description of the submitted lists.

Table 1
Summary of Ranked Dig Lists

ID	Data Used	Description
List 1	EM61 MK2; Metal Mapper	Full target analysis library
List 2	EM61 MK2; Metal Mapper	20mm removed from analysis library
List 3	EM61 MK2; Metal Mapper	Analysis library includes only Camp Butner TOI and clutter
List 4	EM61 MK2; Metal Mapper	Analysis library includes only Camp Butner TOI

Classification by EM61-MK2 Decay Values

Knowing the poor performance of decay-based classification versus Metal Mapper-based classification (see Sibert and SLO), Parsons understood there was little value in seeking a new decay-based approach to compete with recent advances in hardware and software. However, Parsons proceeded with decay-based target classification realizing the value of the discrimination study for comparing decay-based classification approaches. Despite the commercial roll-out of the Metal Mapper and the near ubiquity of the UX-Analyze module, there are situations in which limited resources prevents the use of the latest technologies.

Parameters

Parsons reviewed the 181-item set of training data to identify combinations of decay values for classifying the master list of targets. Decay values were calculated for all combinations of time gates using two different GXs: the UXA Decay.GX, and a GX developed by Parsons. The two .GXs use different methods to determine an anomaly's decay value. The essential difference between the two methods is that Parsons' .GX obtains decay values based on time gate responses at the peak found for each time gate, while the UXA .GX determines the median decay value within a target window. Both .GXs were used to calculate decay values and create ranked dig lists for a comparison of discrimination results. No cued data were used in these dig lists.

UXA Decay.GX data were scatter-plotted in an attempt to find a combination of decay values offering the best separation of TOI from non-TOI. The combination of Tau 1-3 and Tau 2-4 was identified as the most useful in differentiating TOI from non-TOI in the training data (**Figure 2**). Nine dig lists were created using these two decay constants as criteria for target discrimination, four using Parsons' GX for decay calculation (the Peak1 through Peak3, and Peak5), and five using the UXA Decay GX (UXA1 through UXA5). A ninth dig list (Peak4) was created using a combination of Tau 1-2 and Tau 3-4 in an attempt to test at Camp Butner a target classification rule developed during the Camp San Luis Obispo Discrimination Study.

The criteria for target discrimination became more presumptive per dig list. That is, initial iterations for each GX's dig lists are more conservatively based upon the training data, and further iterations progressively reduced the set of Category 2: Can't Decide targets. For brevity, the criteria and results of initial and poorly performing decay-based dig lists are not discussed. Rather, the focus of this section will be upon the several decay-based dig lists that offer insight into the use of decay values alone for target discrimination, and into the differences of the two GXs.

Dig Lists Using Parsons' GX-Calculated Decay Constants

The **Peak4** ranked dig list uses the target classification rule developed during the San Luis Obispo Classification Study. This rule uses parameters Tau 1-2 and Tau 3-4, as calculated using the Parsons .GX, to classify targets based on geometric analyses of training items' feature space. The purpose of the Peak4 ranked dig list is to test the application of a rule developed on a previous project upon the more complex Camp Butner site. Targets are classified "Can't Analyze" in this dig list if any of the following

conditions are met: the target was identified as a clustered target by visual analysis; the quality of the target's peak selection (as found by the Parsons .GX) was inconclusive or open to doubt; the target's EM61-MK2 Channel 2 response is less than 4mV; or the two decay constants used have outlying values (**Table 2**).

Table 2
Categorization Parameters for Peak4 Dig List

		Param1	Param2	other	Clustered targets (263)	trainers (179)	rank ascends by	targets
Peak4 (the SLO rule)	Category 1: Likely Clutter	<2300	>209				descending Param2	255
	Category 2: Can't Decide	<2300	93<x<209				descending Param2	615
	Category 3: Likely TOI	>2300					descending Param1	856
		<2300	<93				ascending Param2	
	Category 4: Can't Analyze			Ch2 < 4mV	x	x	n/a	387
				Ch1_2PTC > 700				
				Ch3_4PTC > 1400				
				Offset MASK = 3				

The **Peak5** ranked dig list uses conservative thresholds of Tau 1-3 and Tau 2-4 as in earlier iterations for categorizing TOI, but uses a small range of decays for Category 2: Can't Decide. Category 4: Can't Analyze targets include targets with null Tau 1-3 or null Tau 2-4. The purpose of the Peak5 ranked dig list is to test the efficacy of the most aggressive target classification criteria for each decay constant calculation method (i.e., Parsons GX or UXA Decay GX) using the same set of Category 4 (**Table 3**).

Table 3
Categorization Parameters for Peak5 Dig List

		Ch1_3PTC	Ch2_4PTC	other	Clustered targets (263)	trainers (179)	rank ascends by	targets
Peak5	Category 1: Likely Clutter	else					ascending CH2_4PTC	938
	Category 2: Can't Decide	360<x<450	560<x<700				ascending CH2_4PTC	735
	Category 3: Likely TOI	>450	>700				ascending CH2_4PTC	168
	Category 4: Can't Analyze			Offset MASK = 3	x	x	n/a	272
				Ch1_2PTC is null				
				Ch2_4PTC is null				

Dig Lists Using UX Analyze-Calculated Decay Constants

The **UXA4** ranked dig list uses a combination of Tau 1-3 and Tau 2-4 based on analyses of training items' and the unclassified targets' parameters. A target is classified as Likely Clutter, Can't Decide, or Likely TOI based on its location in Tau 1-3 and Tau 2-4 feature space (**Figure 4**). Targets are classified "Can't Analyze" in this dig list if: the target was identified as a clustered target by visual analysis, if Tau 1-3 is null, or if Tau 2-4 is null (**Table 4**).

Figure 4

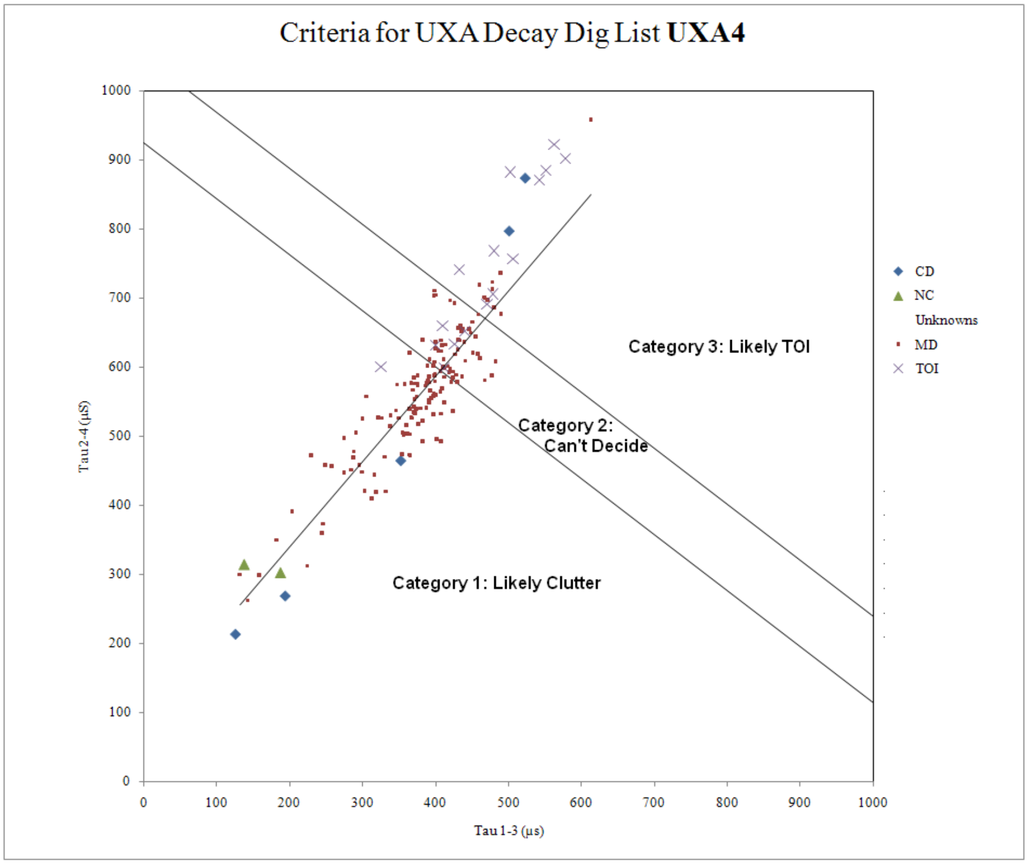


Table 4
Categorization Parameters for UXA4 Dig List

		Tau 1-3	Tau 2-4	other	Clustered targets (263)	trainers (179)	rank ascends by	targets
UXA4	Category 1: Likely Clutter	y = -0.8103x + b, where y = Tau 1-3, and x = Tau 2-4		b<925			ascending b	1065
	Category 2: Can't Decide			925<b<1050			ascending b	557
	Category 3: Likely TOI			b>1050			ascending b	220
	Category 4: Can't Analyze	is null	is null		x	x	n/a	271

The **UXA5** ranked dig list uses conservative thresholds of Tau 1-3 and Tau 2-4 used as in earlier iterations for categorizing TOI, but uses a small range of decays for Category 2: Can't Decide. Category 4: Can't Analyze targets include those meeting any of the criteria in the table below, inclusive of the criteria for Category 4 within the ranked dig list Peak5 (i.e., if the quality of the target's peak selection as found by the Parsons ADV_PROC.GX was inconclusive or open to doubt). The purpose of the UXA5 ranked dig list is to test the efficacy of the most aggressive target classification criteria for each decay constant calculation method (i.e., Parsons' GX or UXA Decay GX) using the same set of Category 4 (**Table 5**).

Table 5
Categorization Parameters for UXA4 Dig List

		Tau 1-3	Tau 2-4	other	Clustered targets (263)	trainers (179)	rank ascends by	targets
UXA5	Category 1: Likely Clutter	else					ascending Tau 2-4	1293
	Category 2: Can't Decide	400<x<475	600<x<700				ascending Tau 2-4	444
	Category 3: Likely TOI	>475	>700				ascending Tau 2-4	104
	Category 4: Can't Analyze			Offset MASK = 3 Tau 1-3 is null Tau 2-4 is null	x	x	n/a	272

Decision Statistic

Decision Statistics for all nine decay-based dig lists were created simply by assigning a value to a target based on its Rank on the dig list: Decision Statistic = 1 – (x/n), where x = the target's Rank and n is the total number of targets. (Category 4: Can't Analyze targets' Decision Statistic is set at -9999).

RESULTS

The submitted dig lists were compared to ground truth data from Camp Butner by the Institute for Defense Analyses upon submittal. Assumptions in this comparison were that targets classified as Category 3, Category 2, and Cannot Analyze would be considered "Dig" targets and that Category 1 targets would be considered "Don't Dig". The results for each list were compared to each other by examining the number of false negatives (TOI incorrectly classified as Category 1), the number of non-TOI digs required to locate at least 95% of the TOI, and the number of digs required to locate 100% of the TOI. These metrics for each of the submitted lists are contained in **Table 2**.

As indicated in **Table 2**, Lists 3 and 4 contained the fewest false negatives with the exception of the Peak4 list (discussed below), both with a total of five. The TOIs identified incorrectly and placed in the "Don't Dig" category were the same for both lists, and there was also little difference between the numbers of digs required to locate 95 percent of the TOI at the site. Categorization for List 3 was significantly more difficult than for List 4 given the necessity of sorting out all of the clutter results returned for targets during advanced target classification. Given the relative ease of dealing with a library without clutter included and lack of significant performance differences, List 4 was deemed to be the most successful for the purposes of this project. **Figure 5** shows the performance curve generated using List 4 and also indicates which TOI were incorrectly identified as "Don't Dig" (blue IDs).

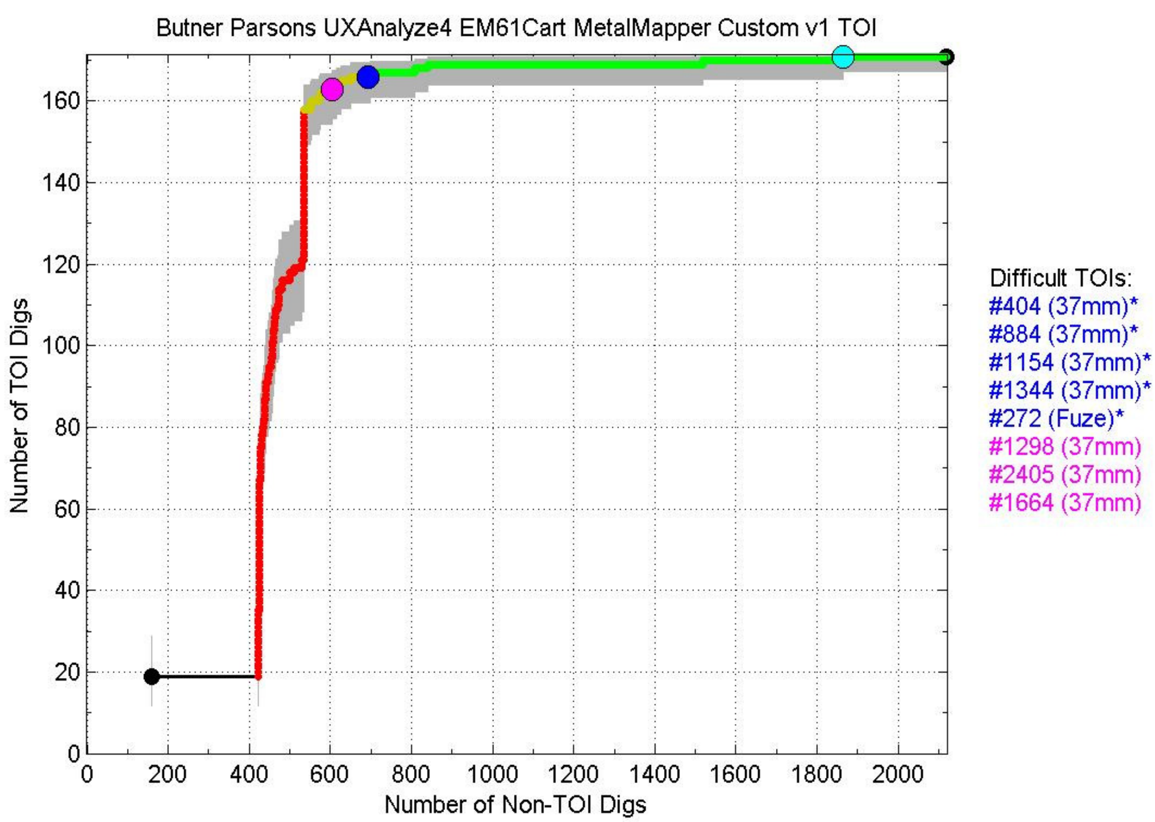
Results for the decay-based dig lists are consistent with expectations. The Peak4 dig list is notable for having zero false negatives. However, this is due to highly exclusive criteria for categorization as non-TOI, which resulted in only 255 targets being classified as non-TOI. Considering the criteria for Category 1 was developed during the Camp San Luis Obispo study using its training data and ordnance items, the zero false negatives at Camp Butner are coincidental, and affirm the effect of conservative thresholds upon false negatives.

Table 2
Ranked Dig List Results

ID	Method	False Negatives	Number of Digs to Locate 95 % of TOI	Number of Digs to Locate 100 % of TOI
List 1	EM61-MK2 decay values and inversion of cued Metal Mapper data	8	1,095	2,155
List 2		10	1,291	2,132
List 3		5	761	2,165
List 4		5	768	2,036
Peak4	EM61-MK2 decay values	0	1,411	1,960
Peak5		11	1,633	2,114
UXA4		12	1,743	2,146
UXA5		44	1,285	2,044

Figure 5

Performance Curve for List 4



FAILURE ANALYSIS

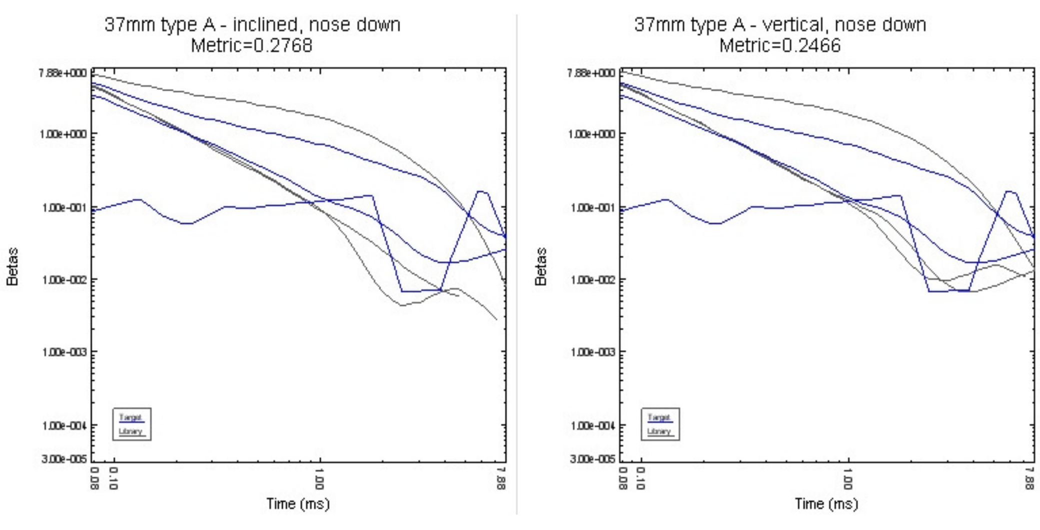
Cued Data: List 4

The following describes the specific reasons that five TOI were incorrectly classified as Category 1, "Don't Dig", targets; and discusses re-analysis options that would possibly results in the re-categorization of these targets as Category 1 or Category 2, "Dig", targets:

- #404 (37mm): Classified based on a tau24 value of 474µs (below 500µs threshold). Metal Mapper data was not requested for this target based on the low tau value. However, it was provided following the project, and analysis resulted in classification as a 37mm with a confidence of 0.794. It is possible that analysis of the Metal Mapper data for this target would have resulted in correct classification as a 37mm.

- #884 (37mm) and #1344 (37mm): Both classified based on a curve matches to TOI, but with confidences below the 0.600 threshold for Category 2. Re-examination of the polarization curves for both indicated that low signal to noise ratios resulted in extremely poor B_3 curves for these targets. **Figure 6** shows the curves generated for #884. Re-analysis options include classification as Category 2 – Can't Analyze based on poor data quality, adding another layer of comparison using only size (B_1) or size and shape 1 (B_2/B_1) to negate the effects of poor B_3 data, or manually examining all of the plots to analyze curve matches.

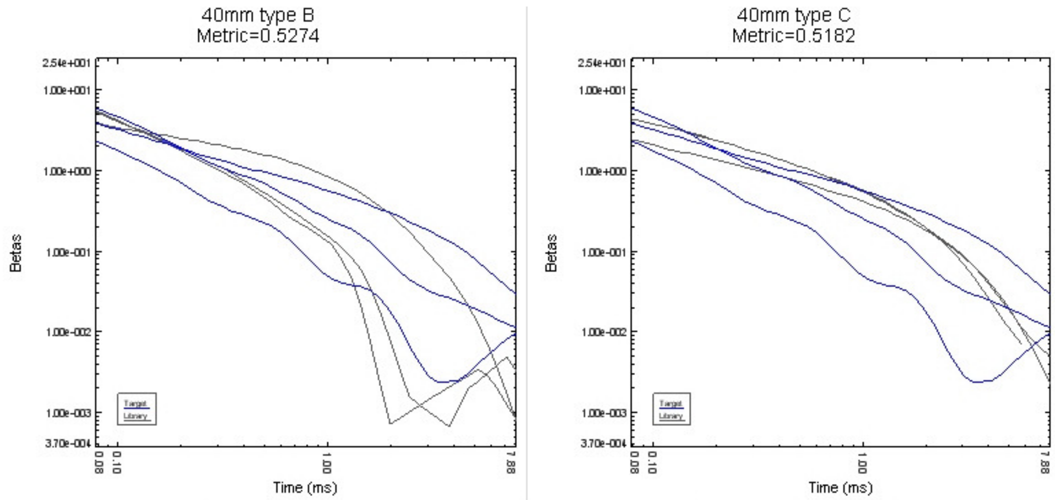
Figure 6
Polarization Curves for Target #884



- #1154 (37mm) and #272 (37mm): Both classified based on a curve matches to TOI, but with confidences below the 0.600 threshold for Category 2. Unlike targets 884 and 1344, the B_3 curves are not obviously unreliable. **Figure 7** shows the curves generated for #1154. Re-analysis options include adding

another layer of comparison using only size (B_1) or size and shape 1 (B_2/B_1) or using a physical examination of the plots to analyze curve matches.

Figure 7
Polarization Curves for Target #1154



EM61-MK2 Data: Peak5 and UXA5 Lists

Decay-based dig lists Peak5 and UXA5, designed to compare the efficacy of decay calculation methods, had unique sets of Difficult TOIs but for similar reasons. In each dig list, Difficult TOIs have decays near thresholds used to classify targets (**Figures 8 and 9**). Although slightly adjusting thresholds would have properly categorized most of these Difficult TOIs, these false negatives illustrate the limitation of the decay property to discriminate TOI.

CONCLUSIONS

The following conclusions were drawn based on the results of the Classification Study performed at the former Camp Butner:

- The EM61-MK2 data alone were not particularly useful for classifying TOI, most likely due to the overlap in size between fragments of larger ordnance (i.e., 105mm projectiles) and the smaller 37mm projectiles at the site.

- The EM data were reasonably useful in identifying decay constant thresholds above and below which targets could be classified as either TOI or non-TOI, respectively. However, the use of the lower threshold to eliminate targets from consideration as TOI did result in the mischaracterization of target #404 as non-TOI. Using EM decay constants saved the collection of approximately 500 Metal Mapper data points. Given the missed TOI and the relatively small amount of data collection saved, the use of EM data to reduce the number of cued targets was unnecessary.

Figure 8

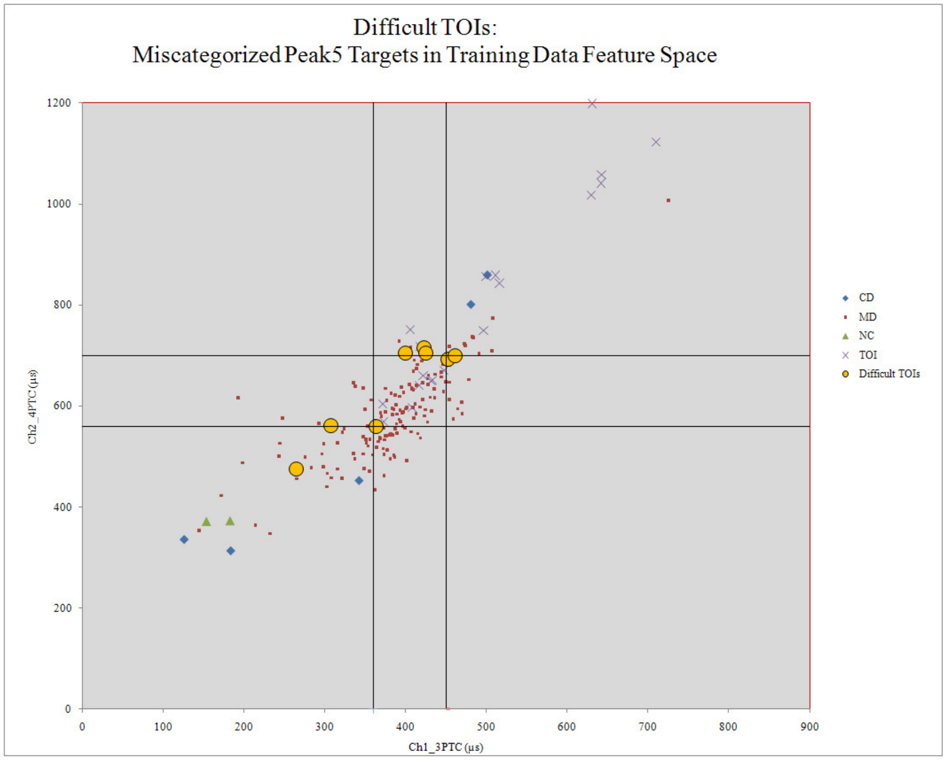
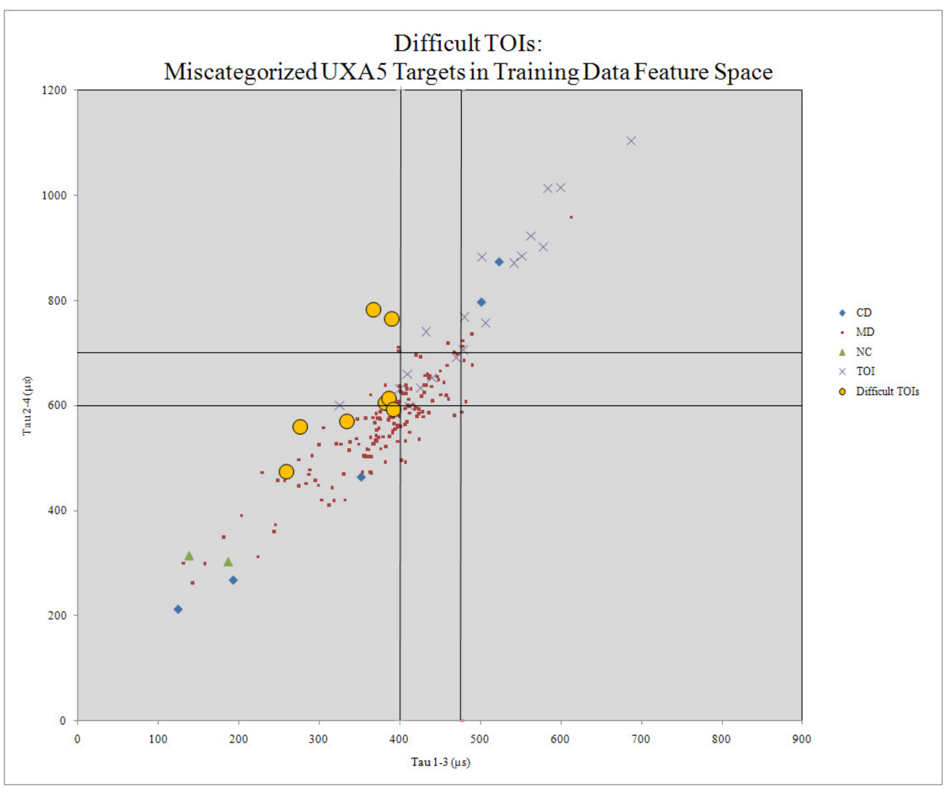


Figure 9



- Removal from the analysis library of TOI not expected at the project site significantly improved both the number of false negatives and the number of digs required to remove 95 percent of the TOI at the site. Furthermore, the removal of clutter from the analysis library did not prove detrimental to classification efforts, and using the Camp Butner TOI-only library was far easier than sorting through a number of results for each target to determine if any were TOI.

- It appears that it would be worthwhile to spend additional analysis time manually examining the curve matches for each target and looking at the use of additional metrics such as the size (B_1) or size and shape 1 (B_2/B_1) when classifying targets to reduce the number of false negatives

- In addition to the conclusions drawn during Parsons' analysis of this data, subsequent discussion of this project and future discrimination projects indicates that the influence of metallic sources in proximity to the target source in question can be effectively separated from the primary object's response using the multiple object solver option in UX-Analyze. Therefore, classifying all clustered targets as Cannot Analyze is unnecessary.

- The UXA5 dig list outperformed its Parsons' GX analog, based on IDA's retrospective 95% and 100% Probability of Detection "Don't Dig" thresholds (i.e., the Pink Dot and Light Blue Dot).